CHAPTER 4

MULTIVARIATE PROBABILITY DISTRIBUTIONS

4.1 Introduction

In Chapters 2 and 3, we studied probability models for a single random variable. Many problems in probability and statistics lead to models involving several random variables simultaneously. For example, a gambler playing blackjack is interested in the event of drawing both an ace and a face card from a 52-card deck. A biologist, observing the number of animals surviving in a litter, is concerned about the intersection of these events:

A: The litter contains n animals.

B: x animals survive.

Similarly, observing both the height and the weight of an individual represents the intersection of a specific pair of events associated with height–weight measurements. In this chapter, we first discuss probability models for the joint behaviour of several random variables, putting special emphasis on the case in which the variables are independent of each other. We then study expected values of functions of several random variables, including covariance and correlation as measures of the degree of association between two variables. Then we will considers conditional distributions, the distributions of random variables given the values of other random variables.

4.2 Bivariate and Multivariate Probability Distributions

There are many experimental situations in which more than one r.v will be of interest to an investigator. We shall first consider joint probability distributions for two discrete r.v's, then for two continuous variables, and finally for more than two variables.

Please enter \offprintinfo{(Title, Edition)}{(Author)}
at the beginning of your document.

4.2.1 Joint Probability Density Function

We already defined the pdf of more that one r.v in Chapter 2 as: Let X_1 and X_2 be two r.vs, then the ***Joint*** pdf of X_1 and X_2 is given by:

$$f(x_1,x_2)=\Pr(X_1=x_1,X_2=x_2)=\begin{cases}\sum\limits_{x_1}\sum\limits_{x_2} f(x_1,x_2), & \text{discrete}\\ \int\limits_{x_1}\int\limits_{x_2} f(x_1,x_2)dx_1dx_2, & \text{continuous}\end{cases}$$

Note: If $f(x_1,x_2,\ldots,x_n)$ is a pdf of r.v's $X_1,X_2,\ldots,X_n$ defined on a sample space $\mathcal{A}=\{(x_1,x_2,\ldots,x_n):-\infty<x_i<\infty; i=1,2,\ldots,n\}$, and if event $A\subset\mathcal{A}$, where $A=\{(x_1,x_2,\ldots,x_n):a_i<x_i<b_i; i=1,2,\ldots,n\}$. Then the probability of A is:

$$p(A)=\Pr\{a_i<x_i<b_i; i=1,2,\ldots,n\}=\begin{cases}\sum\limits_{x_1=a_1}^{b_1}\sum\limits_{x_2=a_2}^{b_2}\cdots\sum\limits_{x_n=a_n}^{b_n} f(x_1,x_2,\ldots,x_n)\\ \\ \int\limits_{x_1=a_1}^{b_1}\int\limits_{x_2=a_2}^{b_2}\cdots\int\limits_{x_n=a_n}^{b_n} f(x_1,x_2,\ldots,x_n)dx_1,dx_2,\ldots,dx_n\end{cases}$$

EXAMPLE 4.1

Let X and Y be discrete r.v's defined on a sample space $\mathcal{A}=\{(x,y):x=1,2,3; y=1,2\}$, and let $f(x,y)$ be a function defined on $\mathcal{A}$ by $f(x,y)=\frac{1}{21}(x+y),\quad (x,y)\in\mathcal{A}$.

1. Is $f(x,y)$ a pdf of X and Y?
2. If so, find $p(A)$ and $p(A^c)$, where $A=\{(x,y):x=1,2; y=1\}$.

Solution.

1. - $f(x,y)>0,\qquad \forall(x,y)\in\mathcal{A}$.
   -
   $$\sum\sum_{(x,y)\in\mathcal{A}} f(x,y)=\sum_{x=1}^{3}\sum_{y=1}^{2}\frac{1}{21}(x+y)=\frac{1}{21}\sum_{x=1}^{3}[(x+1)+(x+2)]$$
   $$=\frac{1}{21}\sum_{x=1}^{3}(2x+3)=\frac{1}{21}[5+7+9]=1.$$

   Then $f(x,y)$ is a pdf of X and Y.

2. $$p(A)=\sum\sum_{(x,y)\in A} f(x,y)=\sum_{x=1}^{2}\sum_{y=1}^{1}\frac{1}{21}(x+y)=\frac{1}{21}\sum_{x=1}^{2}(x+1)=\frac{1}{21}(2+3)=\frac{5}{21}$$
   $$\Rightarrow p(A^c)=1-\frac{5}{21}=\frac{16}{21}.$$

◀

EXAMPLE 4.2

Let X and Y be two r.v's defined on a sample space $\mathcal{A} = \{(x, y) : x = 1, 2, \ldots; y = 0, 1, 2\}$, and let $f(x, y)$ be a function defined on $\mathcal{A}$ by

$$f(x,y) = \binom{2}{y}\left(\frac{1}{2}\right)^{x+2}, \quad (x,y) \in \mathcal{A}$$

1. Is $f(x, y)$ a pdf of X and Y?
2. Find $p(A)$, where $A = \{(x, y) : x = 1, 3, 5, \ldots; y = 1\}$.

Solution.

1. - $f(x, y) > 0, \quad \forall (x, y) \in \mathcal{A}$.

   -
   $$\sum\sum_{(x,y)\in\mathcal{A}} f(x,y) = \sum_{x=1}^{\infty}\sum_{y=0}^{2}\binom{2}{y}\left(\frac{1}{2}\right)^{x+2} = \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^{x+2}\left[\binom{2}{0}+\binom{2}{1}+\binom{2}{2}\right]$$
   $$= \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^{x+2}(1+2+1) = 4\left[\left(\frac{1}{2}\right)^{3}+\left(\frac{1}{2}\right)^{4}+\left(\frac{1}{2}\right)^{5}+\ldots\right] = 4\frac{(1/2)^3}{1-(1/2)} = 1$$

   Then $f(x, y)$ is a pdf of X and Y.

2.
$$p(A) = \sum\sum_{(x,y)\in A} f(x,y) = \sum_{x=1,step2}^{\infty}\sum_{y=1}^{1}\binom{2}{y}\left(\frac{1}{2}\right)^{x+2} = \sum_{x=1,step2}^{\infty}\left(\frac{1}{2}\right)^{x+2}\binom{2}{1} = 2\sum_{x=1,step2}^{\infty}\left(\frac{1}{2}\right)^{x+2}$$
$$= 2\left[\left(\frac{1}{2}\right)^{3}+\left(\frac{1}{2}\right)^{5}+\left(\frac{1}{2}\right)^{7}+\ldots\right] = 2\frac{(1/2)^3}{1-(1/2)^2} = \frac{1}{3}.$$

◀

EXAMPLE 4.3

Let X and Y be continuous r.v's defined on a sample space $\mathcal{A} = \{(x, y) : 0 < x < 2; 2 < y < 4\}$, and let $f(x, y)$ be a function defined on $\mathcal{A}$ as $f(x, y) = \frac{1}{8}(6 - x - y), \quad (x, y) \in \mathcal{A}$

1. Is $f(x, y)$ a pdf of X and Y?
2. Find $p(A)$, where $A = \{(x, y) : 0 < x < 1; 2 < y < 3\}$.

Solution.

1. - $f(x, y) \geq 0, \quad \forall (x, y) \in \mathcal{A}$.

   -
   $$\int\int_{(x,y)\in\mathcal{A}} f(x,y) = \int_{x=0}^{2}\int_{y=2}^{4}\frac{1}{8}(6-x-y)dydx = \frac{1}{8}\int_0^2\left[6y - xy - \frac{y^2}{2}\right]_2^4 dx$$
   $$= \frac{1}{8}\int_0^2\left[(24-4x-8)-(12-2x-2)\right]dx = \frac{1}{8}\int_0^2(6-2x)dx = \frac{1}{8}\left[6x-x^2\right]_0^2 = \frac{12-4}{8} = 1$$

Then $f(x,y)$ is a pdf of X and Y.

2.
$$p(A) = \int\int_{(x,y)\in A} f(x,y) = \int_{x=0}^{1}\int_{y=2}^{3} \frac{1}{8}(6-x-y)dydx = \cdots = \frac{3}{8}$$

◀

EXAMPLE 4.4

X and Y are two r.v's defined on a sample space $\mathcal{A} = \{(x,y) : 0 < x < y < 1\}$, and let $f(x,y) = 2 \quad (x,y) \in \mathcal{A}$ is a function defined on $\mathcal{A}$. is $f(x,y)$ a pdf of X and Y?

Solution.

- $f(x,y) = 2 > 0, \quad \forall (x,y) \in \mathcal{A}$.

-
$$\int\int_{(x,y)\in\mathcal{A}} f(x,y) = 2\int_{x=0}^{1}\int_{y=x}^{1} dydx = 2\int_0^1 y\Big|_x^1 dx = 2\int_0^1 (1-x)dx = 2\left[x - \frac{x^2}{2}\right]_0^1 = 2(1-\frac{1}{2}) = 1$$

Then $f(x,y)$ is a pdf of X and Y. ◀

EXAMPLE 4.5

Find the constant c that makes the function $f(x,y) = ce^{-x-y}, \quad 0 < x < y < \infty$ a pdf of r.v's X and Y.

Solution. Since $f(x,y)$ is a pdf of X and Y, then by definition $\int\int_{(x,y)\in\mathcal{A}} f(x,y)dxdy = 1$:

$$1 = c\int_{x=0}^{\infty}\int_{y=x}^{\infty} e^{-x-y}dydx = c\int_0^\infty e^{-x}\left[\int_x^\infty e^{-y}dy\right]dx = c\int_0^\infty e^{-x}\left[-e^{-y}\right]_x^\infty dx$$

$$= c\int_0^\infty e^{-x}(0+e^{-x})dx = c\int_0^\infty e^{-2x}dx = c\left[-\frac{1}{2}e^{-2x}\right]_0^\infty = c\left(0+\frac{1}{2}\right) = \frac{c}{2}.$$

Therefore, $c = 2$. ◀

EXAMPLE 4.6

A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let X_1 denote the number of customers who choose counter 1 and X_2, the number who select counter 2. Find the joint probability function of X_1 and X_2.

Solution. We might proceed with the derivation in many ways. The most direct is to consider the sample space associated with the experiment. Let the pair $\{i,j\}$ denote the simple event that the first customer chose counter i and the second customer chose counter j , where $i,j = 1,2,$ and 3. Using the mn rule, the sample space consists of $3\times 3 = 9$ sample points. Under the assumptions given earlier, each sample point is equally likely and has probability $1/9$. The sample space associated with the experiment is

$$S = [\{1,1\},\{1,2\},\{1,3\},\{2,1\},\{2,2\},\{2,3\},\{3,1\},\{3,2\},\{3,3\}].$$

Notice that sample point $\{1,1\}$ is the only sample point corresponding to $(X_1 = 2, X_2 = 0)$ and hence $\Pr(X_1 = 2, X_2 = 0) = 1/9$. Similarly, $\Pr(X_1 = 1, X_2 = 1) = \Pr(\{1,2\} \text{ or } \{2,1\}) = 2/9$. Table 5.1 contains the probabilities associated with each possible pair of values for X_1 and X_2—that is, the joint probability function for X_1 and X_2.

Probability function for X_1 and X_2

$x_2 \setminus x_1$	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

◀

EXAMPLE 4.7

A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let $X =$ the proportion of time that the drive-up facility is in use (at least one customer is being served or waiting to be served) and $Y =$ the proportion of time that the walk-up window is in use. Then the set of possible values for (X, Y) is the rectangle $D\{(x,y) : 0 \le x \le 1;\ 0 \le y \le 1\}$. Suppose the joint pdf of (X, Y) is $f(x,y) = \frac{6}{5}(x+y^2), 0 \le x \le 1, 0 \le y \le 1$. Verify that $f(x,y)$ is a valid pdf. Evaluate $\Pr\left(0 \le x \le \frac{1}{4},\ 0 \le Y \le \frac{1}{4}\right)$.

Solution. To verify that this is a legitimate pdf, note that $f(x,y) > 0$ and

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= \int_0^1\int_0^1 \frac{6}{5}(x+y^2)dxdy = \int_0^1\int_0^1 \frac{6}{5}xdxdy + \int_0^1\int_0^1 \frac{6}{5}y^2dxdy \\ &= \int_0^1 \frac{6}{5}xdx(1-0) + \int_0^1 \frac{6}{5}y^2dy(1-0) = \frac{6}{10}x^2\Big|_0^1 + \frac{6}{15}y^3\Big|_0^1 = \frac{6}{10} + \frac{6}{15} = 1\end{aligned}$$

The probability that neither facility is busy more than one-quarter of the time is

$$\begin{aligned}\Pr\left(0 \le X \le \frac{1}{4},\ 0 \le Y \le \frac{1}{4}\right) &= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)dxdy = \frac{6}{5}\int_0^{1/4}\int_0^{1/4} xdxdy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2dxdy \\ &= \frac{6}{5}\left(\frac{1}{4}-0\right)\int_0^{1/4} xdx + \frac{6}{5}\left(\frac{1}{4}-0\right)\int_0^{1/4} y^2dx \\ &= \frac{6}{20}\frac{x^2}{2}\Big|_0^{1/4} + \frac{6}{20}\frac{y^3}{3}\Big|_0^{1/4} = \frac{7}{640} = 0.0109.\end{aligned}$$

◀

EXAMPLE 4.8

Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables take on values between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose that the joint density function for Y_1 and Y_2 is given by: $f(y_1, y_2) = 3y_1, 0 \le y_2 \le y_1 \le 1$. Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

Solution. We want to find $\Pr(0 \leq Y_1 \leq 0.5, Y_2 > 0.25)$. For any continuous random variable, the probability of observing a value in a region is the volume under the density function above the region of interest. We are interested only in values of y_1 and y_2 such that $0 \leq y_1 \leq 0.5$ and $y_2 > 0.25$. Thus we have

$$\begin{aligned}\Pr(0 \leq Y_1 \leq 0.5, Y_2 > 0.25) &= \int_0^{1/2} \int_{1/4}^{y_1} 3y_1 dy_2 dy_1 = \int_0^{1/2} 3y_1 \left(y_2 \Big|_{1/4}^{y_1} \right) dy_1 \\ &= \int_0^{1/2} 3y_1(y_1 - 1/4) dy_1 = \left[y_1^3 - (3/8)y_1^2 \right]_0^{1/2} = 1/32.\end{aligned}$$

◀

4.2.2 Joint Cumulative Distribution Function

Let X_1 and X_2 be any r.v's with joint pdf $f(x_1, x_2)$, then the joint distribution function $F(x_1, x_2)$, such that

$$F(x_1, x_2) = \Pr(X_1 \leq x_1, X_2 \leq x_2) = \begin{cases} \sum\limits_{\tau_1=-\infty}^{x_1} \sum\limits_{\tau_2=-\infty}^{x_2} f(\tau_1, \tau_2), & \text{discrete} \\ \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} f(\tau_1, \tau_2) d\tau_1 d\tau_2, & \text{continuous} \end{cases},$$

for all $-\infty < x_1 < \infty$, $-\infty < x_2 < \infty$.

***Theorem*:** If X_1 and X_2 are r.v's with joint distribution function $F(x_1, x_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. If $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0.$$

EXAMPLE 4.9

Consider the r.v's X_1 and X_2 of Example 4.6. Find $F(-1, 2)$, $F(1.5, 2)$, and $F(5, 7)$.

Solution. Using the results in Example 4.6, we find

$$F(-1, 2) = \Pr(X_1 \leq -1, X_2 \leq 2) = \Pr(\phi) = 0.$$

Further,

$$\begin{aligned}F(1.5, 2) &= \Pr(X_1 \leq 1.5, X_2 \leq 2) \\ &= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(1,2) = 8/9.\end{aligned}$$

Similarly,

$$F(5, 7) = \Pr(X_1 \leq 5, X_2 \leq 7) = 1.$$

Notice that $F(x_1, x_2) = 1$ for all x_1, x_2 such that $\min\{x_1, x_2\} \geq 2$. Also, $F(x_1, x_2) = 0$ if $\min\{x_1, x_2\} < 0$. ◀

EXAMPLE 4.10

Let the pdf of r.v's X and Y be: $f(x,y) = \frac{x}{6}\left(\frac{1}{2}\right)^y;\ x = 1,2,3,\ y = 1,2,\ldots$. Find the cdf of X and Y.

Solution.

$$\begin{aligned} F(x,y) &= \Pr(X \le x, Y \le y) = \sum_{t=-\infty}^{x}\sum_{s=-\infty}^{y} f(t,s) = \sum_{t=1}^{x}\sum_{s=1}^{y}\frac{t}{6}\left(\frac{1}{2}\right)^s \\ &= \sum_{t=1}^{x}\frac{t}{6}\left[\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^y\right] = \sum_{t=1}^{x}\frac{t}{6}\frac{(1/2)[1-(1/2)^y]}{1-(1/2)} \\ &= [1-(1/2)^y]\frac{1}{6}(1+2+\cdots+x) = \frac{1}{6}\frac{x(x+1)}{2}[1-(1/2)^y] \end{aligned}$$

$$\therefore F(x,y) = \begin{cases} 0, & x < 1, y < 1 \\ \frac{x(x+1)}{12}[1-\left(\frac{1}{2}\right)^y], & 1 \le x \le 3, 1 \le y < \infty \\ 1, & x > 3, y = \infty \end{cases}$$

◀

EXAMPLE 4.11

Let X and Y be two r.v's defined on sample space $\mathcal{A} = \{(x,y) : 0 < x < 2;\ 2 < y < 4\}$, and let $f(x,y) = \frac{1}{8}(6-x-y)$ is a pdf of X and Y. Find the cdf.

Solution.

$$\begin{aligned} F(x,y) &= \Pr(X \le x, Y \le y) = \int_{t=-\infty}^{x}\int_{s=-\infty}^{y} f(t,s)dtds = \int_0^x\int_2^y \frac{1}{8}(6-t-s)dsdt \\ &= \frac{1}{8}\int_2^y \left[6t - \frac{1}{2}t^2 - st\right]_0^x ds = \frac{1}{8}\int_2^y \left(6x - \frac{1}{2}x^2 - xs\right)ds = \frac{1}{8}\left[6xs - \frac{1}{2}x^2 s - \frac{1}{2}xs^2\right]_2^y \\ &= \frac{1}{8}\left(6xy - \frac{1}{2}x^2y - \frac{1}{2}xy^2 - 10x + x^2\right) = \frac{x}{16}(12y - xy - y^2 - 20 + 2x). \end{aligned}$$

$$\therefore F(x,y) = \begin{cases} 0, & x \le 0,\ y \le 2 \\ \frac{x}{16}(12y - xy - y^2 - 20 + 2x), & 0 < x < 2,\ 2 < y < 4 \\ 1, & x \ge 2,\ y \ge 4 \end{cases}$$

◀

Note:

1. If $X_1, X_2, \ldots, X_n$ are continuous r.v's, then the pdf is:

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1, \partial x_2, \ldots, \partial x_n}.$$

2. For $2-$dimension discrete sample space, the pdf could be written as:

$$f(x,y) = F(x,y) - F(x,y-1) - F(x-1,y) + F(x-1,y-1).$$

EXAMPLE 4.12

Let the r.v's X, Y and Z have pdf $f(x,y,z) = 6\,e^{-(x+y+z)},\ 0 < x < y < z < \infty$. Find the cdf of X, Y and Z.

Solution.

$$F(x,y,z) = \Pr(X \le x, Y \le y, Z \le z) = \int_{t=-\infty}^{x}\int_{s=-\infty}^{y}\int_{q=-\infty}^{z} f(t,s,q)dtdsdq = 6\int_0^x\int_t^y\int_s^z e^{-(t+s+q)}dqdsdt$$

$$\therefore F(x,y,z) = \begin{cases} 0, & x \le 0,\ y \le 0,\ z \le 0 \\ 6\left[\frac{1}{2}e^{-(x+2y)} - e^{-(x+y+z)} - \frac{1}{6}e^{-3x} + \frac{1}{2}e^{-(2x+z)} - \frac{1}{2}e^{-2y} + e^{-(y+z)} - \frac{1}{2}e^{-z} + \frac{1}{6}\right], & 0 < x < y < z < \infty \\ 1, & x = y = z = \infty \end{cases}$$

◀

4.2.3 Marginal Distribution

Recall that the distinct values assumed by a discrete random variable represent mutually exclusive events. Similarly, for all distinct pairs of values x_1, x_2, the bivariate events $(X_1 = x_1, X_2 = x_2)$, represented by (x_1, x_2), are mutually exclusive events. It follows that the univariate event $(X_1 = x_1)$ is the union of bivariate events of the type $(X_1 = x_1, X_2 = x_2)$, with the union being taken over all possible values for x_2. For example, consider the experiment of tossing a pair of dice. The sample space contains 36 sample points, corresponding to the $mn = (6)(6) = 36$ ways in which numbers may appear on the faces of the dice. Consider the following r.v's

X_1**:** The number of dots appearing on die 1,

X_2**:** The number of dots appearing on die 2.

Then,

$$\begin{aligned}
\Pr(X_1 = 1) &= p(1,1) + p(1,2) + p(1,3) + \cdots + p(1,6) \\
&= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6 \\
\Pr(X_1 = 2) &= p(2,1) + p(2,2) + p(2,3) + \cdots + p(2,6) \\
&= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6 \\
&\vdots \\
\Pr(X_1 = 6) &= p(6,1) + p(6,2) + p(6,3) + \cdots + p(6,6) \\
&= 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6
\end{aligned}$$

Expressed in summation notation, probabilities about the variable X_1 alone are

$$\Pr(X_1 = x_1) = p_1(x_1) = \sum_{x_2=1}^{6} p(x_1, x_2).$$

Similarly, probabilities corresponding to values of the variable X_2 alone are given by

$$\Pr(X_2 = x_2) = p_2(x_2) = \sum_{x_1=1}^{6} p(x_1, x_2).$$

Summation in the discrete case corresponds to integration in the continuous case, which leads us to the following definition.

***Definition*:** Let X_1 and X_2 be jointly r.v's with probability function $f(x_1, x_2)$. Then the marginal probability functions of X_1 and X_2, respectively, are given by

$$f_1(x_1) = \begin{cases} \sum\limits_{x_2} f(x_1, x_2), & discrete \\ \int\limits_{x_2} f(x_1, x_2)dx_2, & continuous \end{cases}$$

$$f_2(x_2) = \begin{cases} \sum\limits_{x_1} f(x_1, x_2), & discrete \\ \int\limits_{x_1} f(x_1, x_2)dx_1, & continuous \end{cases}$$

***Notes*:** Let X and Y are two r.v's with a joint pdf $f(x, y)$ and marginal pdf's $f_X(x)$ and $f_Y(y)$ respectively. Suppose we require to evaluate

1. $\Pr(a \leq X \leq b)$:

$$\Pr(a \leq X \leq b) = \begin{cases} \sum\limits_{x=a}^{b} f_X(x) = \sum\limits_{x=a}^{b} \sum\limits_{y} f(x, y), & discrete \\ \int\limits_{x=a}^{b} f_X(x)dx = \int\limits_{x=a}^{b} \int\limits_{y} f(x, y)dxdy, & continuous \end{cases}$$

2. $\Pr(c \leq Y \leq d)$:

$$\Pr(c \leq Y \leq d) = \begin{cases} \sum\limits_{y=c}^{d} f_Y(y) = \sum\limits_{y=c}^{d} \sum\limits_{x} f(x, y), & discrete \\ \int\limits_{y=c}^{d} f_Y(y)dy = \int\limits_{y=c}^{d} \int\limits_{x} f(x, y)dxdy, & continuous \end{cases}$$

3. $E[u(X)]$:

$$E[u(X)] = \begin{cases} \sum\limits_{x} u(x)f_X(x) = \sum\limits_{x} \sum\limits_{y} u(x)f(x, y), & discrete \\ \int\limits_{x} u(x)f_X(x)dx = \int\limits_{x} \int\limits_{y} u(x)f(x, y)dxdy, & continuous \end{cases}$$

4. $E[u(Y)]$:

$$E[u(Y)] = \begin{cases} \sum\limits_{y} u(y)f_Y(y) = \sum\limits_{y} \sum\limits_{x} u(y)f(x, y), & discrete \\ \int\limits_{y} u(y)f_Y(y)dy = \int\limits_{y} \int\limits_{x} u(y)f(x, y)dxdy, & continuous \end{cases}$$

5. $E[u(X, Y)]$:

$$E[u(X, Y) = \begin{cases} \sum\limits_{y} \sum\limits_{x} u(x, y)f(x, y), & discrete \\ \int\limits_{x} \int\limits_{y} u(x, y)f(x, y)dydx, & continuous \end{cases}$$

EXAMPLE 4.13

Let the joint pdf of r.v's X and Y be:

1. $f(x,y) = \frac{1}{21}(x+y)$, $x = 1,2,3$; $y = 1,2$.
2. $f(x,y) = e^{-(x+y)}$, $0 < x < \infty$; $0 < y < \infty$.

a. Find the marginal pdf for both X and Y.

b. Find $E[X]$, $Var(X)$, $E[Y]$, $Var(Y)$, $E[XY]$.

Solution. For number (1) we have:

a. The marginal pdf of X is

$$f_x(x) = \sum_y f(x,y) = \sum_{y=1}^{2} \frac{1}{21}(x+y) = \frac{1}{21}[(x+1)+(x+2)] = \frac{1}{21}(2x+3), \quad x = 1,2,3.$$

Also, the marginal pdf of Y is

$$f_y(y) = \sum_x f(x,y) = \sum_{x=1}^{3} \frac{1}{21}(x+y) = \frac{1}{21}[(y+1)+(y+2)+(y+3)] = \frac{1}{7}(y+2), \quad y = 1,2.$$

b. To find the expectations,

$$\mu_x = E[X] = \sum_x x f_x(x) = \frac{1}{21}\sum_{x=1}^{3} x(2x+3) = \frac{46}{21}.$$

$$E[X^2] = \sum_x x^2 f_x(x) = \frac{1}{21}\sum_{x=1}^{3} x^2(2x+3) = \frac{114}{21}.$$

$$\sigma_x^2 = Var(X) = E[X^2] - \mu_x^2 = \frac{114}{21} - \left(\frac{46}{21}\right)^2 = \frac{278}{441}.$$

Similarly,

$$\mu_y = E[Y] = \sum_y f_y(y) = \frac{1}{7}\sum_{y=1}^{2} y(y+2) = \frac{11}{7}.$$

$$E[Y^2] = \sum_y y^2 f_y(y) = \frac{1}{7}\sum_{y=1}^{2} y^2(y+2) = \frac{19}{7}.$$

$$\sigma_y^2 = Var(Y) = E[Y^2] - \mu_y^2 = \frac{19}{7} - \left(\frac{11}{7}\right)^2 = \frac{12}{49}.$$

Finally,

$$E[XY] = \sum_x \sum_y xy f(x,y) = \frac{1}{21}\sum_{x=1}^{3}\sum_{y=1}^{2} xy(x+y) = \frac{1}{21}\sum_{x=1}^{3}[x(x+1)+x(x+2)]$$

$$= \frac{1}{21}\sum_{x=1}^{3}(3x^2+5x) = \frac{1}{21}(8+29+42) = \frac{24}{7}.$$

For number (2) we have:

a. The marginal pdf of X is

$$f_x(x) = \int_y f(x,y)dy = \int_0^\infty e^{-(x+y)}dy = -e^{-x}\left[e^{-y}\right]_0^\infty = e^{-x}, \;\; 0 < x < \infty.$$

Also, the marginal pdf of Y is

$$f_y(y) = \int_x f(x,y)dx = \int_0^\infty e^{-(x+y)}dx = -e^{-y}\left[e^{-x}\right]_0^\infty = e^{-y}, \;\; 0 < y < \infty.$$

b. To find the expectations,

$$\mu_x = E[X] = \int_x x f_x(x)dx = \int_0^\infty xe^{-x}dx = 1 = E[Y] = \mu_y.$$

$$E[X^2] = \int_x x^2 f_x(x)dx = \int_0^\infty x^2 e^{-x}dx = 2 = E[Y^2].$$

$$\sigma_x^2 = Var(X) = E[X^2] - \mu_x^2 = 2 - (1)^2 = 1 = \sigma_y^2 = Var(Y)$$

Finally,

$$\begin{aligned} E[XY] &= \int_x \int_y xyf(x,y)dxdy = \int_0^\infty \int_0^\infty xye^{-(x+y)}dxdy \\ &= \left(\int_0^\infty xe^{-x}dx\right)\left(\int_0^\infty ye^{-y}dy\right) = (1)(1) = 1. \end{aligned}$$

◀

***Definition*:** If $f(x,y)$ and $F(x,y)$ are the joint pdf and cdf of r.v's X and Y. Then the marginal cdf of X and Y, respectively , are

$$F_x(x) = F(x,\infty) \;\text{ and }\; F_y(y) = F(\infty,y).$$

and are defined as

$$F_x(x) = \begin{cases} \sum\limits_{\tau=-\infty}^{x} f_x(\tau), & discrete, \\ \int\limits_{-\infty}^{x} f_x(\tau)d\tau, & continuous. \end{cases}$$

and

$$F_y(y) = \begin{cases} \sum\limits_{\tau=-\infty}^{y} f_y(\tau), & discrete, \\ \int\limits_{-\infty}^{y} f_y(\tau)d\tau, & continuous. \end{cases}$$

EXAMPLE 4.14

Back to the last example and find the joint cdf of X and Y, then find the marginal cdf for both X and Y.

Solution. For number (1) we have $f(x,y) = \frac{1}{21}(x+y)$, $x = 1,2,3$; $y = 1,2$. Then

$$F(x,y) = \sum_{\tau=1}^{x}\sum_{\kappa=1}^{y} f(\tau,\kappa) = \frac{1}{21}\sum_{\tau=1}^{x}\sum_{\kappa=1}^{y}(\tau+\kappa) = \frac{1}{21}\sum_{\tau=1}^{x}[(\tau+1)+(\tau+2)+\cdots+(\tau+y)] \tag{4.1}$$
$$= \frac{1}{21}\sum_{\tau=1}^{x}\left[y\tau + \frac{y(y+1)}{2}\right] = \frac{1}{21}\left[\frac{yx(x+1)}{2} + \frac{xy(y+1)}{2}\right] = \frac{1}{42}xy(x+y+2),$$

hence

$$F(x,y) = \begin{cases} 0, & x<1,\ y<1 \\ \frac{1}{42}xy(x+y+2), & 1\le x<3,\ 1\le y<2 \\ 1, & x\ge 3,\ y\ge 2 \end{cases}$$

The marginal cdf of X is

$$F_x(x) = F(x,2) = \begin{cases} 0, & x<1 \\ \frac{1}{21}x(x+4), & 1\le x<3 \\ 1, & x\ge 3 \end{cases}$$

Similarly, the marginal cdf of Y is

$$F_y(y) = F(3,y) = \begin{cases} 0, & y<1 \\ \frac{1}{14}y(y+5), & 1\le y<2 \\ 1, & y\ge 2 \end{cases}$$

For number (2) we have $f(x,y) = e^{-(x+y)}$, $0<x<\infty$; $0<y<\infty$. Then

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(\tau,\kappa)d\tau d\kappa = \int_0^x\int_0^y e^{-(\tau+\kappa)}d\tau d\kappa = \left(\int_0^x e^{-\tau}d\tau\right)\left(\int_0^y e^{-\kappa}d\kappa\right)$$
$$= \left[-e^{-\tau}\right]_0^x\left[-e^{-\kappa}\right]_0^y,$$

hence

$$F(x,y) = \begin{cases} 0, & x\le 0,\ y\le 0 \\ (1-e^{-x})(1-e^{-y}), & 0<x<\inf,\ 0<y<\infty \\ 1, & x=y=\infty \end{cases}$$

The marginal cdf of X is

$$F_x(x) = F(x,\infty) = \begin{cases} 0, & x\le 0 \\ 1-e^{-x}, & 0<x<\infty \\ 1, & x=\infty \end{cases}$$

The marginal cdf of Y is

$$F_y(y) = F(\infty,y) = \begin{cases} 0, & y\le 0 \\ 1-e^{-y}, & 0<y<\infty \\ 1, & y=\infty \end{cases}$$

◀

***Remarks*:**

1. The joint moment generating function (mgf) of the two r.v's X and Y with a joint pdf $f(x,y)$ is defined as

$$M(t_1,t_2) = E[e^{t_1X+t_2Y}] = \begin{cases} \sum\limits_x \sum\limits_y e^{t_1X+t_2Y} f(x,y), & discrete \\ \int\limits_x \int\limits_y e^{t_1X+t_2Y} f(x,y)dxdy, & continuous \end{cases}$$

   and satisfy $M(0,0) = 1$.

2. The marginal mgf of X and Y can be obtained by

$$M_X(t_1) = E[e^{t_1X}] = M(t_1,0), \quad \text{and} \quad M_Y(t_2) = E[e^{t_2Y}] = M(0,t_2).$$

3. We can evaluate the mean and variance for both X and Y using the mgf as

$$E[X^nY^m] = \left.\frac{\partial^{n+m} M(t_1,t_2)}{\partial t_1^n \partial t_2^m}\right|_{t_1=t_2=0}$$

4.2.4 Conditional Distribution

The distribution of Y can depend strongly on the value of another variable X. For example, if X is height and Y is weight, the distribution of weight for men who are 180 cm tall is very different from the distribution of weight for short men. The conditional distribution of Y given $X = x$ describes for each possible x how probability is distributed over the set of possible y values. We define the conditional distribution of Y given X, but the conditional distribution of X given Y can be obtained by just reversing the roles of X and Y. Both definitions are analogous to that of the conditional probability $\Pr(A|B)$ as the ratio $\Pr(A \cap B)/\Pr(B)$.

***Definition*:** If X and Y are jointly r.v's with joint density function $f(x,y)$, then the conditional pdf of X given $Y = y$ is

$$f_{X|Y}(x) = \Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} = \frac{f(x,y)}{f_Y(y)},$$

also, conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y) = \Pr(Y = y|X = x) = \frac{\Pr(Y = y, X = x)}{\Pr(X = x)} = \frac{f(x,y)}{f_X(x)}.$$

***Definition*:** If X and Y are jointly r.v's with joint density function $f(x,y)$, then the conditional distribution function of X given $Y = y$ and Y given $X = x$ are

$$F(x|y) = \Pr(X \le x|Y = y), \quad \text{and} \quad F(y|x) = \Pr(Y \le y|X = x).$$

EXAMPLE 4.15

Let the joint pdf of r.v's X and Y be $f(x,y) = e^{-y}$, $0 < x < y < \infty$. Find the following

1. The conditional pdf of both X given Y and Y given X.
2. $E[X|Y = y]$, $Var(X|Y = y)$, $E[Y|X = x]$ and $Var[Y|X = x]$.
3. Evaluate the probabilities: $\Pr(1 \le X \le 2)$ and $\Pr(1 \le X \le 2|Y = 2)$.

Solution. We have the joint pdf $f(x,y) = e^{-y}$, $0 < x < y < \infty$, then the marginal pdf of X and Y are

$$f_X(x) = \int_y f(x,y)dy = \int_x^\infty e^{-y}dy = -e^{-y}\Big|_x^\infty = e^{-x}, \quad 0 < x < \infty.$$

and

$$f_Y(y) = \int_x f(x,y)dx = \int_0^y e^{-y}dy = e^{-y}x\Big|_0^y = ye^{-y}, \quad 0 < y < \infty.$$

1. The conditional pdf of X given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}, \quad 0 < x < y, \text{ for any } 0 < y < \infty.$$

   The conditional pdf of Y given $X = x$ is

$$f_2(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \quad x < y < \infty, \text{ for any } 0 < x < \infty.$$

2. To evaluate the expectations:

$$E[X|Y=y] = \mu_{x|y} = \int_x x f_1(x|y)dx = \int_0^y x\frac{1}{y}dx = \frac{1}{y}\left[\frac{x^2}{2}\right]_0^y = \frac{y}{2}.$$

$$E[X^2|Y=y] = \int_x x^2 f_1(x|y)dx = \int_0^y x^2\frac{1}{y}dx = \frac{1}{y}\left[\frac{x^3}{3}\right]_0^y = \frac{y^2}{3}.$$

$$\therefore \sigma^2_{x|y} = E[X^2|Y=y] - \mu^2_{x|y} = \frac{y^2}{3} - \left(\frac{y}{2}\right)^2 = \frac{y^2}{12}.$$

   Similarly,

$$E[Y|X=x] = \mu_{y|x} = \int_y y f_2(y|x)dy = \int_x^\infty ye^{-(y-x)}dx = e^x\int_x^\infty ye^{-y}dy$$
$$= e^x\left[-ye^{-y} - e^{-y}\right]_x^\infty = e^x(xe^{-x} + e^{-x}) = x + 1.$$

$$E[Y^2|X=x] = \int_y y^2 f_2(y|x)dy = \int_x^\infty y^2e^{-(y-x)}dx = e^x\int_x^\infty y^2e^{-y}dy$$
$$= e^x\left[-y^2e^{-y} - 2ye^{-y} - 2e^{-y}\right]_x^\infty = e^x(x^2e^{-x} + 2xe^{-x} + 2e^{-x}) = x^2 + 2x + 2.$$

$$\sigma_{y|x} = E[Y^2|X=x] - \mu^2_{y|x} = x^2 + 2x + 2 - (x+1)^2 = 1.$$

3. To Evaluate the probabilities:

$$\Pr(1 \le X \le 2) = \int_1^2 f_X(x)dx = \int_1^2 e^{-x}dx = -e^{-x}\Big|_1^2 = e^{-1} - e^{-2} = 0.233.$$

$$\Pr(1 \le X \le 2|Y=2) = \int_1^2 f_1(x|y=2)dx = \int_1^2 \frac{1}{2}dx = \frac{x}{2}\Big|_1^2 = \frac{1}{2}(2-1) = \frac{1}{2}.$$

◀

EXAMPLE 4.16

A soft-drink machine has a random amount Y in supply at the beginning of a given day and dispenses a random amount X during the day (with measurements in gallons). It is not resupplied during the day, and hence $X \leq Y$. It has been observed that X and Y have a joint density function $f(x, y) = \frac{1}{2}$, $0 \leq x \leq y \leq 2$. That is, the points (x, y) are uniformly distributed over the triangle with the given boundaries. Find the conditional density of X given $Y = y$. Evaluate the probability that less than 1/2 gallon will be sold, given that the machine contains 1.5 gallons at the start of the day.

Solution. The marginal density of Y is given by

$$f_Y(y) = \int_x f(x, y)dx = \int_0^y \frac{1}{2}dx = \frac{y}{2}, \quad 0 \leq y \leq 2.$$

Note that $f_Y(y) > 0$ if and only if $0 < y \leq 2$. Thus, for any $0 < y \leq 2$,

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{1/2}{y/2} = \frac{1}{y}, \quad 0 \leq x \leq y.$$

Also, $f(x|y)$ is undefined if $y \leq 0$ or $y > 2$. The probability of interest is

$$\Pr(X \leq 1/2|Y = 1.5) = \int_0^{1/2} f(x|y = 1.5)dx = \int_0^{1/2} \frac{1}{1.5}dx = \frac{1/2}{3/2} = \frac{1}{3}.$$

◀

EXAMPLE 4.17

Let the r.v X has a pdf $f(x)$ and cdf $F(x)$. Define $f(x|X > x_0) = \frac{f(x)}{1-F(x_0)}$, $x_0 < x < \infty$ and x_0 is a fixed number.

1. Show that $f(x|X > x_0)$ be a conditional pdf of X given $X > x_0$.
2. Consider $f(x) = e^{-x}$, $0 < x < \infty$. Compute $\Pr(X > 2|X > 1)$.

Solution. 1. To prove that $f(x|X > x_0)$ is a pdf, we check

- $f(x|X > x_0) \geq 0$, because $f(x) \geq 0$ and $1 - F(x_0) > 0$.
- The unity of integration,

$$\int_{x_0}^{\infty} f(x|X > x_0)dx = \int_{x_0}^{\infty} \frac{f(x)}{1 - F(x_0)}dx = \frac{1}{1 - F(x_0)} \int_{x_0}^{\infty} f(x)dx = \frac{1}{1 - F(x_0)} [F(x)]_{x_0}^{\infty}$$
$$= \frac{F(\infty) - F(x_0)}{1 - F(x_0)} = \frac{1 - F(x_0)}{1 - F(x_0)} = 1.$$

2. Since we have the pdf $f(x) = e^{-x}$ and the cdf is defined as: $F(x) = 1 - e^{-x}$ (exponential distribution). Therefore,

$$f(x|X > 1) = \frac{f(x)}{1 - F(1)} = \frac{e^{-x}}{1 - 1 + e^{-1}} = e \cdot e^{-x}, \quad 1 < x < \infty.$$

Now, to compute the probability

$$\Pr(X > 2|X > 1) = \int_2^{\infty} f(x|X > 1)dx = \int_2^{\infty} e \cdot e^{-x}dx = -e \left[e^{-x}\right]_2^{\infty} = -e(0 - e^{-2}) = e^{-1}.$$

◀

4.2.5 Independent Random Variables

In many situations, information about the observed value of one of the two variables X and Y gives information about the value of the other variable. In Example 4.16 we saw two dependent r.v's, for which probabilities associated with X depended on the observed value of Y. In other examples, this is not always the case. Probabilities associated with X were the same, regardless of the observed value of Y. We now present a formal definition of independence of r.v's. Two events A and B are independent if $\Pr(A \cup B) = \Pr(A) \times \Pr(B)$.

***Definition*:** Two random variables X and Y with joint density function $f(x, y)$ and marginal density functions $f_X(x)$ and $f_Y(y)$, are said to be independent if for every pair of x and y values,

$$f(x, y) = f_X(x) \cdot f_Y(y), \quad \forall (x, y).$$

we can also define the independence in terms of the joint cdf and the marginal cdf's $F_X(x)$ and $F_Y(y)$ of X and Y, then

$$F(x, y) = F_X(x) \cdot F_Y(y),$$

for every pair of real numbers (x, y).

For the die-tossing problem of Section 4.2.3, show that X_1 and X_2 are independent since each of the 36 sample points was given probability 1/36. Consider, for example, the point (1, 2). We know that $p(1, 2) = 1/36$. Also, $p_1(1) = \Pr(X_1 = 1) = 1/6$ and $p_2(2) = \Pr(X_2 = 2) = 1/6$. Hence, $p(1, 2) = p_1(1)p_2(2)$. The same is true for all other values for x_1 and x_2, and it follows that X_1 and X_2 are independent.

***Note*:** If two r.v's X and Y are independent, we have

$$f_1(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x).$$

and

$$f_2(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y).$$

EXAMPLE 4.18

Back to example 4.13 and examine whether or not X and Y are independent.

Solution. In example 4.13 we have two cases:

1. The first case where X and Y are two discrete r.v's with joint pdf $f(x, y) = \frac{1}{21}(x + y)$, $x = 1, 2, 3$; $y = 1, 2$. and marginal pdf's $f_X(x) = \frac{1}{21}(2x + 3)$, $x = 1, 2, 3$ and $f_Y(y) = \frac{1}{7}(y + 2)$, $y = 1, 2$. In order to check if X and Y are independent,
$$f_X(x) \cdot f_Y(y) = \frac{1}{217}(2x + 3)(y + 2) \neq \frac{1}{21}(x + y) = f(x, y).$$
Therefore X and Y are dependent r.v's.

2. The second case where X and Y are two continuous r.v's with joint pdf $f(x, y) = e^{-(x+y)}$, $0 < x, y < \infty$, and marginal pdf's $f_X(x) = e^{-x}$, $0 < x < \infty$ and $f_Y(y) = e^{-y}$, $0 < y < \infty$. In order to check if X and Y are independent,
$$f_X(x) \cdot f_Y(y) = e^{-x} \cdot e^{-y} = e^{-(x+y)} = f(x, y),$$
this implies that X and Y are independent.

◀

***Properties*:** Let X and Y are two independent r.v's, then:

1. $\Pr(a \leq X \leq b, c \leq Y \leq d) = \Pr(a \leq X \leq b) \cdot \Pr(c \leq Y \leq d)$.
2. $E[u(X) \cdot v(Y)] = E[u(X)] \cdot E[v(Y)]$.
3. $M(t_1, t_2) = M_X(t_1) \cdot M_Y(t_2)$.

EXAMPLE 4.19

Let the joint mgf of r.v's X and Y is: $M(t_1, t_2) = \frac{e^{\frac{t_1^2}{1-2t_2}}}{1-2t_2}$. Find the marginal mgf's for both X and Y and test whether X and Y are independent.

Solution. We can simply find the marginal mgf's for X and Y as

$$M_X(t_1) = M(t_1, 0) = e^{t_1^2}, \quad \text{and} \quad M_Y(t_2) = M(0, t_2) = (1 - 2t_2)^{-1}.$$

In order to test if X and Y are independent,

$$M_X(t_1) \cdot M_Y(t_2) = \frac{e^{t_1^2}}{1 - 2t_2} \neq M(t_1, t_2),$$

therefore X and Y are dependent. ◄

***theorem*:** Let X and Y have a joint density $f(x, y)$, Then X and Y are independent r.v's if and only if

$$f(x, y) = g(x)h(y)$$

where $g(x)$ is a non-negative function of x alone and $h(y)$ is a non-negative function of y alone.

EXAMPLE 4.20

Let X and Y have a joint density given by: $f(x, y) = 2x$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Are X and Y independent variables?

Solution. Notice that $f(x, y)$ is positive for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Further, $f(x, y) = g(x)h(y)$, where

$$g(x) = x, \quad 0 \leq x \leq 1, \quad \text{and} \quad h(y) = 2, 0 \leq y \leq 1.$$

Therefore, X and Y are independent r.v's. Notice that $g(x)$ and $h(y)$, as defined here, are not density functions, although $2g(x)$ and $h(y2)/2$ are densities. ◄

4.3 Expected Values, Covariance, and Correlation

We previously saw that any function $u(X)$ of a single r.v X is itself a r.v. However, to compute $E[u(X)]$, it was not necessary to obtain the probability distribution of $u(X)$; instead, $E[u(X)]$ was computed as a weighted average of $u(X)$ values, where the weight function was the pdf $f(x)$ of X. A similar result holds for a function $u(X, Y)$ of two jointly distributed r.v's.

EXAMPLE 4.21

Let X and Y have joint density given by: $f(x,y) = 2x$, $0 \leq x \leq 1$; $0 \leq y \leq 1$. Find $E[XY]$, $E[X]$ and $Var(X)$.

Solution. From expectation definition we obtain

- The expected value of $E[XY]$ is:

$$E[XY] = \int_0^1 \int_0^1 xyf(x,y)dxdy = \int_0^1 \int_0^1 xy(2x)dxdy = \int_0^1 y\left[\frac{2x^3}{3}\right]_0^1 dy$$
$$= \frac{2}{3}\int_0^1 ydy = \frac{2}{3}\left[\frac{y^2}{2}\right]_0^1 = \frac{1}{3}.$$

- The expected value of $E[X]$ is:

$$E[X] = \int_0^1 \int_0^1 xf(x,y)dxdy = \int_0^1 \int_0^1 x(2x)dxdy = \int_0^1 \left[\frac{2x^3}{3}\right]_0^1 dy$$
$$= \frac{2}{3}\int_0^1 dy = \frac{2}{3}[y]_0^1 = \frac{2}{3}.$$

- To find $Var(X)$ we need to find $E[X^2]$ first, then:

$$E[X^2] = \int_0^1 \int_0^1 x^2 f(x,y)dxdy = \int_0^1 \int_0^1 x^2(2x)dxdy = \int_0^1 \left[\frac{2x^4}{4}\right]_0^1 dy$$
$$= \frac{1}{2}\int_0^1 dy = \frac{1}{2}[y]_0^1 = \frac{1}{2}.$$

Therefore,

$$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

◀

EXAMPLE 4.22

A process for producing an industrial chemical yields a product containing two types of impurities. For a specified sample from this process, let X denote the proportion of impurities in the sample and let Y denote the proportion of type I impurities among all impurities found. Suppose that the joint distribution of X and Y can be modelled by the pdf: $f(x,y) = 2(1-x)$, $0 \leq x \leq 1$; $0 \leq y \leq 1$. Find the expected value of the proportion of type I impurities in the sample.

Solution. Because X is the proportion of impurities in the sample and Y is the proportion of type I impurities among the sample impurities, it follows that XY is the proportion of type I impurities in the entire sample. Thus, we want to find $E[XY]$:

$$E[XY] = \int_0^1 \int_0^1 2xy(1-x)dxdy = 2\int_0^1 x(1-x)\left(\frac{1}{2}\right)dx = \int_0^1 (x-x^2)dx$$
$$= \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

Therefore, we would expect 1/6 of the sample to be made up of type I impurities. ◀

4.3.1 The Covariance and Correlation Coefficient

Intuitively,we think of the dependence of two random variables X and Y as implying that one variable, say X, either increases or decreases as Y_2 changes. We will confine our attention to two measures of dependence: the covariance between two random variables and their correlation coefficient.

4.3.1.1 The Covariance When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to each other. The average value of $(X-\mu_x)(Y-\mu_y)$ provides a measure of the linear dependence between X and Y. This quantity, $E[(X-\mu_x)(Y-\mu_y)]$, is called the covariance of X and Y.

***Definition*:** If X and Y are r.v's with means μ_x and μ_y and variances σ_x^2 and σ_y^2, respectively, the covariance of X and Y is defined as

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)].$$

The larger the absolute value of the covariance of X and Y, the greater the linear dependence between X and Y. Positive values indicate that X increases as Y increases; negative values indicate that X decreases as Y increases. A zero value of the covariance indicates that the variables are uncorrelated and that X and Y are independent.

***Note*:** From the definition of the covariance, we can find the relation between the variances and covariance of the two r.v's X and Y as the following:

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] = E[XY-\mu_xY-\mu_yX+\mu_x\mu_y] = E[XY]-\mu_xE[Y]-\mu_yE[X]+\mu_x\mu_y \\ &= E[XY]-\mu_x\mu_y-\mu_y\mu_x+\mu_x\mu_y = E[XY]-\mu_x\mu_y \end{aligned}$$

EXAMPLE 4.23

Let the joint pdf of r.v's X and Y is defined as $f(x,y) = x+y$, $0<x<1$, $0<y<1$. Compute the covariance of X and Y.

Solution. In order to evaluate the covariance of X and Y, we need to calculate first μ_x and μ_y. So lets start with computing the marginal pdf's for both X and Y. Then,

$$f_X(x) = \int_y f(x,y)dy = \int_0^1 (x+y)dy = \left[xy+\frac{1}{2}y^2\right]_0^1 = x+\frac{1}{2}, \quad 0<x<1.$$

and

$$f_Y(y) = y+\frac{1}{2}, \quad 0<y<1.$$

Secondly, we have to evaluate μ_x and μ_y, thus

$$\mu_x = E[X] = \int_x x f_X(x)dx = \int_0^1 x\,(x+1/2)\,dx = \left[\frac{x^3}{3}+\frac{x^2}{4}\right]_0^1 = \frac{7}{12}.$$

Similarly, $\mu_y = \frac{7}{12}$. Now the expected value of XY is calculated by

$$\begin{aligned} E[XY] &= \int_x\int_y xyf(x,y)dxdy = \int_0^1\int_0^1 xy(x+y)dxdy = \int_0^1\int_0^1 (x^2y+xy^2)dxdy \\ &= \int_0^1 \left[\frac{x^2y^2}{2}+\frac{xy^3}{3}\right]_0^1 dx = \int_0^1 \left(\frac{x^2}{2}+\frac{x}{3}\right)dx = \left[\frac{x^3}{6}+\frac{x^2}{6}\right]_0^1 = \frac{1}{3}. \end{aligned}$$

Therefore, the covariance of X and Y is

$$Cov(X,Y) = E[XY]-\mu_x\mu_y = \frac{1}{3}-\left(\frac{7}{12}\right)\left(\frac{7}{12}\right) = -\frac{1}{144}.$$

◀

EXAMPLE 4.24

A nut company markets cans of deluxe mixed nuts containing almonds, cashews, and peanuts. Suppose the net weight of each can is exactly 1 lb, but the weight contribution of each type of nut is random. Because the three weights sum to 1, a joint probability model for any two gives all necessary information about the weight of the third type. Let $X =$ the weight of almonds in a selected can and $Y =$ the weight of cashews. Then the region of positive density is $D = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1; x + y \le 1\}$. Now let the joint pdf for (X, Y) be $f(x, y) = 24xy$, $0 \le x \le 1$, $0 \le y \le 1$; $x + y \le 1$. Find the probability that the two types of nuts together make up at most 50% of the can. Find The marginal pdf for almonds. Evaluate the expected amount of almonds and of cashews at each can. Then find the covariance of X and Y.

Solution. - Let $A = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1; x + y \le 0.5\}$, and we need to calculate the probability of event A, that is

$$\Pr\{(X, Y) \in A\} = \iint_A f(x, y)dxdy = \int_0^{0.5} \int_0^{0.5-x} 24xydydx = 0.0625.$$

- The marginal pdf for almonds is obtained by holding X fixed at x and integrating $f(x, y)$ along the vertical line through x:

$$f_X(x) = \int_0^{1-x} f(x, y)dy = \int_0^{1-x} 24xydy = 12x(1 - x)^2, \quad 0 \le x \le 1.$$

- The expected amount of almond at each can is obtained as:

$$\mu_x = E[X] = \int_0^1 x f_X(x)dx = \int_0^1 12x^2(1 - x)^2 dx = \frac{2}{5}.$$

Similarly, the expected amount of cashews at these cans can be easily evaluated by by replacing x by y in $f_X(x)$. Then we get $\mu_y = \frac{2}{5}$.

- Finally, to examine how those two r.v's are related to each other, we calculate the covariance of X and Y. Then

$$E[XY] = \int_0^1 \int_0^{1-x} xy24xydydx = \frac{2}{15}.$$

Thus,

$$Cov(X, Y) = E[XY] - \mu_x\mu_y = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}.$$

A negative covariance is reasonable here because more almonds in the can implies fewer cashews.

◀

***Note*:** If X, Y and Z are r.v's and a and b are constants, then

$$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z).$$

4.3.1.2 Correlation Coefficient It is difficult to employ the covariance as an absolute measure of dependence because its value depends upon the scale of measurement. This problem can be eliminated by standardizing its value and using the correlation coefficient, ρ, a quantity related to the covariance and defined as

$$\rho = \frac{Cov(X, Y)}{\sigma_x\sigma_y}.$$

where σ_x and σ_y are the standard deviations of X and Y, respectively. The sign of the correlation coefficient is the same as the sign of the covariance. Thus, $\rho > 0$ indicates that Y increases as X increases, and $\rho = +1$ implies perfect correlation, with all points falling on a straight line with positive slope. A value of $\rho = 0$ implies zero covariance and no correlation. A negative coefficient of correlation implies a decrease in Y as X increases, and $\rho = -1$ implies perfect correlation, with all points falling on a straight line with negative slope.

EXAMPLE 4.25

From example 4.23, find the correlation coefficient.

Solution. To calculate the correlation coefficient, we need to calculate the standard deviation for both X and Y, and therefore $E[X^2]$ and $E[Y^2]$, then

$$E[X^2] = \int_x x^2 f_X(x)dx = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx = \left[\frac{x^4}{4} + \frac{x^3}{6}\right]_0^1 = \frac{5}{12}.$$

Hence

$$\sigma_x^2 = E[X^2] - \mu_x^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Similarly, $\sigma_y^2 = \frac{11}{144}$. Therefore,

$$\rho = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}} = -\frac{1}{11}.$$

◀

***Theorem*:** If X and Y are independent r.v's, the $\rho = 0$. However, if $\rho = 0$ does not imply independence.

EXAMPLE 4.26

Let X and Y be discrete r.v's with joint probability distribution as shown in the table below. Show that X and Y are dependent but have zero covariance.

Probability function for X and Y

$y \setminus x$	-1	0	1
-1	1/16	3/16	1/16
0	3/16	0	3/16
1	1/16	3/16	1/16

Solution. Calculating the marginal pdf's yields, $f_X(-1) = f_X(1) = 5/16 = f_Y(-1) = f_Y(1)$, and $f_X(0) = 6/16 = f_Y(0)$. Then, we obtain the following marginal pdf's for both X and Y:

Marginal pdf of X and Y

x	$f_X(x)$		y	$f_Y(y)$
-1	5/16		-1	5/16
0	6/16		0	6/16
1	5/16		1	5/16

The value $f(0,0) = 0$ in the centre cell stands out. Obviously, $f(0,0) \neq f_X(0)f_Y(0)$, and this is sufficient to show that X and Y are dependent.

Again looking at the marginal probabilities, we see that $E[X] = E[Y] = 0$. Also,

$$\begin{aligned} E[XY] &= \sum_x \sum_y xyf(x,y) \\ &= (-1)(0)(3/16) + (-1)(1)(1/16) + (0)(-1)(3/16) \\ &+ (0)(0)(0) + (0)(1)(3/16) + (1)(-1)(1/16) + (1)(0)(3/16) + (1)(1)(1/16) \\ &= (1/16) - (1/16) - (1/16) + (1/16) = 0 \end{aligned}$$

Thus,

$$Cov(X, y) = E[XY] - E[X]E[Y] = 0 - (0)(0) = 0$$

This example shows that if the covariance of two r.v's is zero, the variables need not be independent. ◄

EXAMPLE 4.27

Annie and Alvie have agreed to meet for lunch between noon (0:00 p.m.) and 1:00 p.m. Denote Annie's arrival time by X, Alvie's by Y, and suppose X and Y are independent with pdf's:

$$f_X(x) = 3x^2, \quad 0 \le x \le 1: \quad \text{and} \quad f_Y(y) = 2y, \quad 0 \le y \le 1.$$

What is the expected amount of time that the one who arrives first must wait for the other person?

Solution. First of all, the joint pdf of X and Y is found by:

$$f(x, y) = f_X(x) \cdot f_Y(y) = 3x^2 \times 2y = 6x^2y$$

Assuming that X and Y are the arrival time of Annie and Alive respectively give as that the waiting time of the one who comes first is $u(X, Y) = |X - Y|$, then the expected waiting time is $E[|X - Y|]$, and since

$$u(X, Y) = |X - Y| = \begin{cases} X - Y, & x - y \ge 0 \Rightarrow x \ge y \Rightarrow 0 \le y \le x \le 1, \\ Y - X, & x - y < 0 \Rightarrow x < y \Rightarrow 0 < x < y < 1, \end{cases}$$

and then

$$\begin{aligned} E[|X - Y|] &= \int_0^1 \int_0^x (x - y)(6x^2y)dydx + \int_0^1 \int_0^y (y - x)(6x^2y)dxdy \\ &= \frac{1}{6} + \frac{1}{12} = \frac{1}{4}. \end{aligned}$$

Therefore, the expected waiting time for anyone who comes first is 1/4 of an hour. ◄

CHAPTER 5

FUNCTION OF RANDOM VARIABLES

5.1 Introduction

In Chapter 2 we discussed the problem of starting with a single r.v X, forming some function of X, such as X^2 or e^X, to obtain a new r.v $Y = h(X)$, and investigating the distribution of this new r.v. We now generalize this scenario by starting with more than a single r.v. Consider as an example a system having a component that can be replaced just once before the system itself expires. Let X_1 denote the lifetime of the original component and X_2 the lifetime of the replacement component. Then any of the following functions of X_1 and X_2 may be of interest to an investigator:

1. The total lifetime $X_1 + X_2$.

2. The ratio of lifetimes X_1/X_2; for example, if the value of this ratio is 2, the original component lasted twice as long as its replacement.

3. The ratio $X_1/(X_1 + X_2)$, which represents the proportion of system lifetime during which the original component operated.

To determine the probability distribution for a function of n random variables, $X_1, X_2, \ldots, X_n$, we must find the joint probability distribution for the r.v's themselves. We generally assume that observations are obtained through random sampling. We will assume throughout the remainder of this text that populations are large in comparison to the sample size and consequently that the random variables obtained through a random sample are in fact independent of one another. We will present three methods for finding the probability distribution for a function of random variables.

Consider r.v's $X_1, X_2, \ldots, X_n$ and a function $U(X_1, X_2, \ldots, X_n)$, denoted simply as U. We will present two method to find the probability function of the function of r.v's.

Please enter \offprintinfo{(Title, Edition)}{(Author)}
at the beginning of your document.

5.2 The Joint pdf of a Function of Random Variables

Given set of r.v's $X_1, X_2, \ldots, X_n$, consider forming new r.v's $Y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, \ldots, k$, $k \leq n$.

$$f(x_1, x_2, \ldots, x_n) = \text{ the joint pdf of the original r.v's,}$$
$$g(y_1, y_2, \ldots, y_k) = \text{ the joint pdf of the new r.v's.}$$

We now focus on finding the joint distribution of these new variables. For our next study, we will use the term random sample which is defined as:

***Definition*: (Random Sample):** Let $X_1, X_2, \ldots, X_n$ be independent r.v's each of which has the same pdf $f(x)$, that is $f_i(x_i) = f(x_i)$, $i = 1, 2, \ldots, n$, then the joint pdf of these r.v's $f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$. Hence, the r.v's $X_1, X_2, \ldots, X_n$ are said to constitute a random sample (r.s) of size n from a distribution $F(x)$.

For instant, if the two r.v's $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent with joint pdf

$$f(x, y) = \prod_{i=1}^{2} f_i(x_i) = \left[\frac{1}{\sqrt{2\pi}\sigma_1^2} e^{-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2}\right]\left[\frac{1}{\sqrt{2\pi}\sigma_2^2} e^{-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2}\right].$$

Therefore, X_1 and X_2 are said to be a r.s of size two.

Definition*:** A function of one or more r.v's which is not depend on any unknown parameter is called a ***Statistic. For example, $Y = \sum X_i$ is a statistic, while $Y = X - \mu/\sigma$ is not a statistic unless μ and σ is known. A statistic is any quantity whose value can be calculated from sample data. Therefore, a statistic is a random variable and will be denoted by an upper case letter; a lower case letter is used to represent the calculated or observed value of the statistic.

***Definitions*:** Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from any distribution, we define

1. The statistic $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean.
2. The statistic $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(X_i^2 - n\bar{X}\right)\right]$ is called the sample variance.

5.2.1 The Method of Transformations

We will illustrate the method of transformations in two cases. The first is when the r.v's $X_1, X_2, \ldots, X_n$ are discrete, and the second case for continuous set of r.v's $X_1, X_2, \ldots, X_n$.

5.2.1.1 Discrete r.v's Let $X_1, X_2, \ldots, X_n$ be n r.v's of discrete type defined on the n-dimensional space $\mathcal{A}$ with joint pdf $f(x_1, x_2, \ldots, x_n)$. Let the r.v's $Y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, k$ and $k \leq n$ be functions of $X_1, X_2, \ldots, X_n$. The joint pdf of $Y_1, Y_2, \ldots, Y_k$ is required. If $k < n$, we will introduce additional new r.v's $Y_{k+1} = u_{k+1}(X_1, X_2, \ldots, X_n)$, $Y_{k+2} = u_{k+2}(X_1, X_2, \ldots, X_n)$, $\ldots Y_n = u_n(X_1, X_2, \ldots, X_n)$, so that the function $y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, n$ are defined to be one-to-one transformation that maps the space $\mathcal{A}$ of $\{X_i\}$ onto the space $\mathcal{B}$ of $\{Y_i\}$ with inverse transforms $x_i = \omega_i(y_1, y_2, \ldots, y_n)$, $i = 1, 2, \ldots, n$. Then the joint pdf of $Y_1, Y_2, \ldots, Y_n$ is defined as:

$$g(y_1, y_2, \ldots, y_n) = f(\omega_1(y_1, y_2, \ldots, y_n), \omega_2(y_1, y_2, \ldots, y_n), \ldots, \omega_n(y_1, y_2, \ldots, y_n)).$$

Therefore, the joint pdf of $Y_1, Y_2, \ldots, Y_k$ is

$$g^*(y_1, y_2, \ldots, y_k) = \sum_{y_{k+1}} \sum_{y_{k+2}} \cdots \sum_{y_n} g(y_1, y_2, \ldots, y_n).$$

We can also evaluate the marginal pdf of each r.v Y_i using the definition of the marginal density function.

EXAMPLE 5.1

Let the r.v X has a pdf $f(x) = \frac{x}{15}$, $x = 1, 2, 3, 4, 5$. Find the pdf of the r.v $Y = 2X + 1$.

Solution. The function $y = 2x + 1$ defines a one-to-one transformation that maps the space $\mathcal{A} = \{x : x = 1, 2, 3, 4, 5\}$ onto the space $\mathcal{B} = \{y : y = 3, 5, 7, 9, 11\}$ with inverse $x = \frac{1}{2}(y - 1)$. Then the pdf of Y is

$$g(y) = f\left(\frac{y-1}{2}\right) = \frac{1}{30}(y-1), \quad y = 3, 5, 7, 9, 11.$$

◀

EXAMPLE 5.2

Let the r.v $X \sim b(n, p)$. What is the distribution of r.v $Y = n - X$?

Solution. The pdf of X is given by: $f(x) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, \ldots, n$. The function $y = n - x$ defines a one-to-one transform that maps the space $\mathcal{A} = \{x : x = 0, 1, \ldots, n\}$ onto the space $\mathcal{B} = \{y : y = 0, 1, \ldots, n\}$ with inverse function of $x = n - y$. Then the pdf of Y is

$$g(y) = f(n-y) = \binom{n}{n-y} p^{n-y} q^{n-(n-y)} = \binom{n}{y} q^y p^{n-y}, \quad y = 0, 1, \ldots, n.$$

Hence, the r.v $Y \sim b(n, q)$. ◀

EXAMPLE 5.3

Suppose that X_1 and X_2 are two independent r.v's where $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$ (i.e X_1 and X_2 are a r.s of size 2). Find the distribution of the r.v $Y = X_1 + X_2$.

Solution. Since X_1 and X_2 are a r.s, then the joint pdf of X_1 and X_2 can be determined as

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!} = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{x_2}}{x_1!x_2!}, \quad x_i = 0, 1, 2, \ldots;\ i = 1, 2.$$

◀

Consider, beside $Y = X_1 + X_2$, $Z = X_2$. The functions $y = x_1 + x_2$ and $z = x_2$ define one-to-one transformation that maps the space $\mathcal{A} = \{(x_1, x_2) : x_i = 0, 1, 2, \ldots;\ i = 1, 2\}$ onto the space $\mathcal{B} = \{(y, z) : 0 \le z \le y \le \infty\}$, with inverse functions $x_1 = y - z$ and $x_2 = z$. Then the joint pdf of Y and Z is

$$g(y, z) = f(y - z, z) = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{y-z}\lambda_2^z}{(y-z)!z!}, \quad (y, z) \in \mathcal{B}.$$

Now, the marginal pdf of Y is obtained by

$$g_Y(y) = \sum_z g(y, z) = e^{-(\lambda_1+\lambda_2)} \sum_{z=0}^{y} \frac{\lambda_1^{y-z}\lambda_2^z}{(y-z)!z!} = \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \sum_{z=0}^{y} \frac{y!}{(y-z)!z!} \lambda_1^{y-z}\lambda_2^z$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \sum_{z=0}^{y} \binom{y}{z} \lambda_2^z \lambda_1^{y-z} = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^y}{y!}, \quad y = 0, 1, 2, \ldots.$$

Therefore $Y \sim P(\lambda_1 + \lambda_2)$.

***Note*:** We can prove that if $X_1, X_2, \ldots, X_n$ are r.s of size n and $X_i \sim P(\lambda_i)$, then $Y = \sum_{i=1}^n X_i \sim P(\sum_{i=1}^n \lambda_i)$.

EXAMPLE 5.4

Let X_1 and X_2 be independent r.v's of discrete type with joint pdf $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$, where $f_i(x_i)$ is the marginal pdf of X_i, $i = 1, 2$. Consider $Y_i = u_i(X_i)$, $i = 1, 2$ define one-to-one transformation from $\mathcal{A}$ onto $\mathcal{B}$. Show that the r.v's $Y_1 = u_1(X_1)$ and $Y_2 = u_2(X_2)$ are also independent.

Solution. Given $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ where $f(x_1, x_2)$ is the joint pdf of X_1 and X_2 with $f_i(x_i)$ is the marginal pdf of X_i, $i = 1, 2$.

Also the functions $y_i = u_i(x_i)$, $i = 1, 2$ maps $\mathcal{A}$ onto $\mathcal{B}$, then the inverse functions $x_i = \omega_i(y_i)$, $i = 1, 2$. Then the joint pdf of Y_1 and Y_2 is:

$$g(y_1, y_2) = f(\omega_1(y_1), \omega_2(y_2)) = f_1(\omega_1(y_1)) \cdot f_2(\omega_2(y_2)).$$

◀

Then, the marginal pdf of Y_1 is obtained by

$$g_1(y_1) = \sum_{y_2} g(y_1, y_2) = f_1(\omega_1(y_1)) \sum_{y_2} f_2(\omega_2(y_2)) = c_1 f_1(\omega_1(y_1)),$$

similarly, the marginal pdf of Y_2 is

$$g_2(y_2) = \sum_{y_1} g(y_1, y_2) = f_2(\omega_2(y_2)) \sum_{y_1} f_1(\omega_1(y_1)) = c_2 f_2(\omega_2(y_2)).$$

In order to verify the independence of Y_1 and Y_2, we need to prove that $g(y_1, y_2) = g_1(y_1) \cdot g_2(y_2)$. Therefore

$$\begin{aligned} g_1(y_1) \cdot g_2(y_2) &= c_1 f_1(\omega_1(y_1)) \cdot c_2 f_2(\omega_2(y_2)) = c_1 c_2 f_1(\omega_1(y_1)) \cdot f_2(\omega_2(y_2)) \\ &= c_1 c_2 g(y_1, y_2), \\ \Rightarrow g(y_1, y_2) &= \frac{1}{c_1 c_2} g_1(y_1) \cdot g_2(y_2). \end{aligned}$$

Since $g(y_1, y_2)$ is the joint pdf of Y_1 and Y_2, then

$$\begin{aligned} 1 &= \sum_{y_1} \sum_{y_2} g(y_1, y_2) = \sum_{y_1} \sum_{y_2} \frac{1}{c_1 c_2} g_1(y_1) \cdot g_2(y_2) = \frac{1}{c_1 c_2} \left[\sum_{y_1} g_1(y_1) \right] \left[\sum_{y_2} g_2(y_2) \right] \\ &= \frac{1}{c_1 c_2} \Rightarrow c_1 c_2 = 1 \\ \therefore g(y_1, y_2) &= g_1(y_1) \cdot g_2(y_2). \end{aligned}$$

This implies that Y_1 and Y_2 are independent r.v's.

5.2.1.2 Continuous r.v's In the case of continuous r.v's we assume the same environment as it was with the discrete case. Let $X_1, X_2, \ldots, X_n$ be n r.v's of continuous type defined on the n-dimensional space $\mathcal{A}$ with joint pdf $f(x_1, x_2, \ldots, x_n)$. Let the r.v's $Y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, k$ ($k \leq n$) be functions of $X_1, X_2, \ldots, X_n$. The joint pdf of $Y_1, Y_2, \ldots, Y_k$ is required. If $k < n$, we will introduce additional new r.v's $Y_{k+1} = u_{k+1}(X_1, X_2, \ldots, X_n)$, $Y_{k+2} = u_{k+2}(X_1, X_2, \ldots, X_n)$, $dots Y_n = u_n(X_1, X_2, \ldots, X_n)$, so that the function $y_i = u_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2, \ldots, n$ are defined to be one-to-one transformation that maps the space $\mathcal{A}$ of $\{X_i\}$ onto the space $\mathcal{B}$ of $\{Y_i\}$ with inverse transforms $x_i = \omega_i(y_1, y_2, \ldots, y_n)$, $i = 1, 2, \ldots, n$. In this case, we will need to calculate the Jacobian $n \times n$ determinant of the first partial derivatives as:

$$J = \frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(y_1, y_2, \ldots, y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Then the joint pdf of $Y_1, Y_2, \ldots, Y_n$ is defined as:

$$g(y_1, y_2, \ldots, y_n) = |J| \times f(\omega_1(y_1, y_2, \ldots, y_n), \omega_2(y_1, y_2, \ldots, y_n), \ldots, \omega_n(y_1, y_2, \ldots, y_n)),$$

and finally, the joint pdf of $Y_1, Y_2, \ldots, Y_k$ is

$$g^*(y_1, y_2, \ldots, y_k) = \int_{y_{k+1}} \int_{y_{k+2}} \cdots \int_{y_n} g(y_1, y_2, \ldots, y_n) dy_{k+1} dy_{k+2} \ldots dy_n.$$

We can also evaluate the marginal pdf of each r.v Y_i using the definition of the marginal density function.

EXAMPLE 5.5

Let the r.v X has a pdf $f(x) = 3x^2$, $0 < x < 1$. Find the pdf of the r.v $Y = X^3$ and evaluate $\Pr(\frac{1}{2} < Y < \frac{3}{4})$.

Solution. The function $y = x^3$ is a one-to-one function from $\mathcal{A} = \{x : 0 < x < 1\}$ onto $\mathcal{B} = \{y : 0 < y < 1\}$ with inverse function $x = y^{1/3}$. The Jacobian in the 1-dimensional case is obtained as

$$J = \frac{\partial x}{\partial y} = \frac{1}{3} y^{-2/3}.$$

Therefore, the pdf of the r.v Y is

$$g(y) = f(y^{1/3})|J| = 3(y^{1/3})^2 \frac{1}{3} y^{-2/3} = 1, \quad 0 < y < 1.$$

Now, to evaluate the probability $\Pr(\frac{1}{2} < Y < \frac{3}{4})$,

$$\Pr\left(\frac{1}{2} < Y < \frac{3}{4}\right) = \int_{1/2}^{3/4} dy = y\Big|_{1/2}^{3/4} = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}.$$

◀

EXAMPLE 5.6

Let the r.v $X \sim U(0, 1)$. Find and name the distribution of the r.v $Y = -2 \ln X$.

Solution. The function $y = -2 \ln x$ is a one-to-one function that maps $\mathcal{A} = \{x : 0 < x < 1\}$ onto $\mathcal{B} = \{y : 0 < y < \infty\}$, with inverse function $x = e^{-\frac{1}{2}y}$. The Jacobian,

$$J = \frac{\partial x}{\partial y} = -\frac{1}{2} e^{-\frac{1}{2}y}.$$

Then, the pdf of Y can be obtained as

$$g(y) = |J| f(e^{-\frac{1}{2}y}) = (1)\frac{1}{2} e^{-\frac{1}{2}y} = \frac{1}{2} e^{-\frac{1}{2}y}, \quad 0 < y < \infty$$

Hence, the r.v $Y \sim Exp(2)$. ◀

EXAMPLE 5.7

Let X_1 and X_2 be a r.s of size 2 from $N(0,1)$. Define the r.v's $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint pdf of Y_1 and Y_2 and show that they are independent.

Solution. The joint pdf of X_1 and X_2 is:

$$f(x_1,x_2) = f_1(x_1) \times f_2(x_2) = \frac{1}{2\pi} e^{-\frac{1}{2\pi}(X_1^2+X_2^2)}, \quad -\infty < x_i < \infty;\ i = 1,2.$$

The functions $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$ define as one-to-one functions that map the space $\mathcal{A} = \{(x_1,x_2) : -\infty < x_i < \infty;\ i = 1,2\}$ onto the space $\mathcal{B} = \{(y_1,y_2) : -\infty < y_i < \infty;\ i = 1,2\}$, with inverses $x_1 = \frac{1}{2}(y_1 + y_2)$ and $x_2 = \frac{1}{2}(y_1 - y_2)$. The Jacobian can be calculated as

$$J = \frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}.$$

Therefore, the joint pdf of y_1 and y_2 is

$$g(y_1,y_2) = f\left(\frac{1}{2}(y_1+y_2), \frac{1}{2}(y_1-y_2)\right) \times |J|.$$

taking the term $x_1^2 + x_2^2$:

$$x_1^2 + x_2^2 = \frac{1}{4}(y_1+y_2)^2 + \frac{1}{4}(y_1-y_2)^2 = \frac{1}{4}\left[y_1^2 + 2y_1y_2 + y_2^2 + y_1^2 - 2y_1y_2 + y_2^2\right] = \frac{1}{4}(2y_1^2 + 2y_2^2) = \frac{1}{2}(y_1^2 + y_2^2).$$

Therefore,

$$g(y_1,y_2) = \frac{1}{4\pi} e^{-\frac{1}{4}(y_1^2+y_2^2)}, \quad -\infty < y_i < \infty;\ i = 1,2.$$

To verify if Y_1 and Y_2 are independent or not, we need to find the marginal pdf's for both Y_1 and Y_2. Then, the marginal pdf of Y_1 is

$$\begin{aligned} g_1(y_1) &= \int_{y_2} g(y_1,y_2)dy_2 = \frac{1}{4\pi} e^{-\frac{1}{4}y_1^2} \int_{-\infty}^{\infty} e^{-\frac{1}{4}y_2^2} dy_2 = \frac{\sqrt{2\pi}\sqrt{2}}{4\pi} e^{-\frac{1}{4}y_1^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{4}y_2^2} dy_2 \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_1^2}{2}} \quad -\infty < y_1 < \infty. \end{aligned}$$

That is $Y_1 \sim N(0,2)$. Similarly, the marginal pdf of Y_2 is $g_2(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_2^2}{2}}$, $-\infty < y_2 < \infty$, which means that $Y_2 \sim N(0,2)$. Then,

$$g_1(y_1) \cdot g_2(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_1^2}{2}} \times \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_2^2}{2}} = \frac{1}{4\pi} e^{-\frac{1}{4}(y_1^2+y_2^2)} = g(y_1,y_2).$$

Therefore, Y_1 and Y_2 are independent r.v's. ◀

EXAMPLE 5.8

Consider $n = 3$ identical components with independent lifetimes X_1, X_2, X_3, each having an exponential distribution with parameter $1/\lambda$. If the first component is used until it fails, replaced by the second one which remains in service until it fails, and finally the third component is used until failure, then the total lifetime of these components is $Y = X_1 + X_2 + X_3$. Find the distribution of total lifetime.

Solution. The joint pdf of X_1, X_2 and X_3 is $f(x_1, x_2, x_3) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)}$, $0 < x_i < \infty$, $i = 1, 2, 3$. We define two other new variables $Y_1 = X_1$ and $Y_2 = X_1 + X_2$ (so that $Y_1 < Y_2 < Y$). Then the functions $y = x_1 + x_2 + x_3$, $y_2 = x_1 + x_2$ and $y_1 = x_1$ define a one-tp-one transformation that maps $\mathcal{A} = \{(x_1, x_2, x_3) : 0 < x_i < \infty, i = 1, 2, 3\}$ onto the space $\mathcal{B} = \{(y_1, y_2, y_3) : 0 < y_1 < y_2 < y < \infty\}$, with inverses:

$$x_1 = y_1, \quad x_2 = y_2 - y_1 \quad \text{and} \quad x_3 = y - y_2.$$

The Jacobian is calculated as

$$J = \frac{\partial(x_1, x_2, x_3)}{\partial(y_1, y_2, y)} = \begin{vmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{vmatrix} = 1$$

Then, the joint pdf of Y_1, Y_2 and Y is

$$g(y_1, y_2, y) = f(y_1, y_2 - y_1, y - y_2)|J| = \lambda^3 e^{-\lambda y}, \quad 0 < y_1 < y_2 < y < \infty.$$

Finally, the marginal pdf of Y is

$$\begin{aligned} g^*(y) &= \int_{y_1}\int_{y_2} g(y_1, y_2, y)dy_1dy_2 = \int_0^y \int_{y_1}^y \lambda^3 e^{-\lambda y} dy_1 dy_2 = \int_0^y \lambda^3 e^{-\lambda y}\Big[y_2\Big]_{y_1}^y dy_1 \\ &= \int_0^y \lambda^3 e^{-\lambda y}(y - y_1)dy_1 = \lambda^3 e^{-\lambda y}\left[yy_1 - \frac{y_1^2}{2}\right]_0^y = \lambda^3 e^{-\lambda y}\left[y^2 - \frac{y^2}{2}\right] \\ &= \frac{\lambda^3}{2} y^2 e^{-\lambda y}, \quad 0 < y < \infty. \end{aligned}$$

◀

5.3 Order Statistics

Many functions of random variables of interest in practice depend on the relative magnitudes of the observed variables. For instance, we may be interested in the fastest time in an automobile race or the heaviest mouse among those fed on a certain diet. Thus, we often order observed random variables according to their magnitudes. The resulting ordered variables are called order statistics.

Formally, let $X_1, X_2, \ldots, X_n$ denote independent continuous random variables with distribution function $F(x)$ and density function $f(x)$. We denote the ordered random variables X_i by $Y_1, Y_2, \ldots, Y_n$, where $Y_1 \le Y_2 \le \cdots \le Y_n$. where

$$\begin{aligned} Y_1 &= \text{the smallest among } X_1, X_2, \ldots, X_n \\ Y_2 &= \text{the second smallest among } X_1, X_2, \ldots, X_n \\ &\vdots \\ Y_n &= \text{the largest among } X_1, X_2, \ldots, X_n \end{aligned}$$

i.e, $Y_1 = \min\{X_1, X_2, \ldots, X_n\}$ and $Y_n = \max\{X_1, X_2, \ldots, X_n\}$.

***Notes*:**

1. In most applications, the unordered set of r.v's $\{X_i\}$ is usually taken to be a r.s of size n.

2. We only consider the continuous type of r.v's.

5.3.1 The distribution of Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from a continuous distribution having a pdf of $f(x)$ and cdf $F(x)$, where the joint pdf of those r.v's is

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i), \quad -\infty < x_i < \infty;\ i = 1, 2, \ldots, n.$$

Suppose that $Y_1, Y_2, \ldots, Y_n$ is a ordered sample of the original r.s, that is $Y_1 = \min\{X_1, X_2, \ldots, X_n\}$ and $Y_n = \max\{X_1, X_2, \ldots, X_n\}$. The probability density functions for Y_1 and Y_n can be found using the method of distribution functions. We will derive the density function of Y_n first. Because Y_n is the maximum of $X_1, X_2, \ldots, X_n$, the event $(Y_n \leq y)$ will occur if and only if the events $(X_i \leq y)$ occur for every $i = 1, 2, \ldots, n$. That is,

$$\Pr(Y_n \leq y) = \Pr(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y).$$

Because the X_i are independent and $\Pr(X_i \leq y) = F(y)$ for $i = 1, 2, \ldots, n$, it follows that the distribution function of Y_n is given by

$$F_{Y_n}(y) = \Pr(Y_n \leq y) = \Pr(X_1 \leq y) \Pr(X_2 \leq y) \ldots \Pr(X_n \leq y) = [F(y)]^n.$$

Letting $g_n(y)$ denote the density function of Y_n, we see that, on taking derivatives of both sides,

$$g_n(y) = n[F(y)]^{n-1} f(y).$$

The density function for Y_1 can be found in a similar manner. The distribution function of Y_1 is

$$F_{Y_1}(y) = \Pr(Y_1 \leq y) = 1 - \Pr(Y_1 > y).$$

Because Y_1 is the minimum of $X_1, X_2, \ldots, X_n$, it follows that the event $(Y_1 > y)$ occurs if and only if the events $(X_i > y)$ occur for $i = 1, 2, \ldots, n$. Because the X_i are independent and $\Pr(Y_i > y) = 1 - F(y)$ for $i = 1, 2, \ldots, n$, we see that

$$\begin{aligned} F_{Y_1}(y) &= \Pr(Y_1 \leq y) = 1 - \Pr(Y_1 > y) \\ &= 1 - \Pr(X_1 > y, X_2 > y, \ldots, X_n > y) \\ &= 1 - [\Pr(X_1 > y) \Pr(X_2 > y) \ldots \Pr(X_n > y)] \qquad = 1 - [1 - F(y)]^n. \end{aligned}$$

Thus, if $g_1(y)$ denotes the density function of Y_1, differentiation of both sides of the last expression yields

$$g_1(y) = n[1 - F(y)]^{n-1} f(y).$$

In general, The marginal pdf of the kth order statistic Y_k could be derived as

$$g_k(y) = \frac{n!}{(k-1)!(n-k)!} [F(y)]^{k-1} [1 - F(y)]^{n-k} f(y), \quad -\infty < y < \infty.$$

EXAMPLE 5.9

Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a r.s of size 4 from a distribution having a pdf of $f(x) = 2x$, $0 < x < 1$. Compute $\Pr(Y_3 > 1/2)$ and $E[Y_3]$.

Solution. Since the pdf of the r.s is $f(x) = 2x$, $0 < x < 1$. Then we can find the cdf as

$$F(x) = \int_{-\infty}^{x} f(\tau)d\tau = \int_0^x 2\tau d\tau = \tau\Big|_0^x = x^2.$$

Now, since $n = 4$ and $k = 3$, then the pdf of Y_3 is

$$g_3(y) = \frac{4!}{2!1!}y^4(1-y^2)2y = 24y^5(1-y^2), \quad 0 < y < 1.$$

Hence, the probability

$$\Pr\left(Y_3 > \frac{1}{2}\right) = \int_{1/2}^{1} g_3(y)dy = \int_{1/2}^{1} 24y^5(1-y^2)dy = 24\left[\frac{y^6}{6} - \frac{y^8}{8}\right]_{1/2}^{1} = 0.949.$$

The mean value of Y_3 is

$$E[Y_3] = \int_0^1 yg_3(y)dy = 24\int_0^1 y^6(1-y^2)dy = 24\left[\frac{y^7}{7} - \frac{y^9}{9}\right]_0^1 = \frac{16}{21}.$$

◀

EXAMPLE 5.10

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from $Exp(\lambda)$. Find the distribution of the first order statistics Y_1.

Solution. Since $X \sim Exp(\lambda)$, then $f(x) = \frac{1}{\lambda}e^{-x/\lambda}$, $0 < x < \infty$, and the cdf $F(x) = 1 - e^{-x/\lambda}$. Therefore, the pdf of Y_1 is given by

$$g_1(y) = n[1-F(y)]^{n-1}f(y) = n\left[1 - 1 + e^{-\frac{y}{\lambda}}\right]^{n-1}\frac{1}{\lambda}e^{-\frac{y}{\lambda}} = \frac{n}{\lambda}\left[e^{-\frac{y}{\lambda}}\right]^{n-1}e^{-\frac{y}{\lambda}} = \frac{n}{\lambda}e^{-\frac{n}{\lambda}y}, \quad 0 < y < \infty.$$

◀

EXAMPLE 5.11

Electronic components of a certain type have a length of life X, with probability density given by: $f(x) = (1/100)e^{-x/100}$, $x > 0$. (Length of life is measured in hours.) Suppose that two such components operate independently and in series in a certain system (hence, the system fails when either component fails). Find the density function for Y, the length of life of the system.

Solution. Because the system fails at the first component failure, $Y = \min\{X_1, X_2\}$, where X_1 and X_2 are independent r.v's with the given density. Then, because $F(x) = 1 - e^{-x/100}$, for $x \geq 0$,

$$g_1(y) = n[1-F(y)]^{n-1}f(y) = 2e^{-\frac{y}{100}}\frac{e^{-\frac{y}{100}}}{100} = \frac{1}{50}e^{-\frac{y}{50}}, \quad y > 0.$$

Thus, the minimum of two exponentially distributed random variables has an exponential distribution. Notice that the mean length of life for each component is 100 hours, whereas the mean length of life for the system is $E[Y] = 50 = 100/2$. ◀

EXAMPLE 5.12

Let X denote the contents of a one-gallon container, and suppose that its pdf is $f(x) = 2x$ for $0 < x < 1$ (and 0 otherwise) with corresponding cdf $F(x) = x^2$ in the interval of positive density. Consider a random sample of four such containers. Determine the expected value of $Y_4 - Y_1$, the difference between the contents of the most-filled container and the least-filled container.

Solution. To determine $Y_4 - Y_1$, the sample range, The pdf's of Y_4 and Y_1 are

$$g_4(y) = n[F(y)]^{n-1} f(y) = 4(y^2)^3 \cdot 2y = 8y^7, \quad 0 \le y \le 1,$$

and

$$g_1(y) = n[1 - F(y)]^{n-1} f(y) = 4(1 - y^2)^3 \cdot 2y = 8y(1 - y^2)^3, \quad 0 \le y \le 1.$$

Therefore, the expected value of $Y_4 - Y_1$ is

$$E[Y_4 - Y_1] = E[Y_4] - E[Y_1] = \int_0^1 y \cdot 8y^7 dy - \int_0^1 y \cdot 8y(1 - y^2)^3 dy = 0.889 - 0.406 = 0.483.$$

If random samples of four containers were repeatedly selected and the sample range of contents determined for each one, the long run average value of the range would be 0.483. ◀

5.3.2 Joint Distribution of the Order Statistics

Consider a r.s $X_1, X_2, \ldots, X_n$ of size n with a joint pdf

$$f(x_1, x_2, \ldots, x_n) = f(x_1) \cdot f(x_2) \ldots f(x_n),$$

and $f(x_i)$ is the marginal pdf of each r.v X_i. Then joint pdf of r.v's $Y_1, Y_2, \ldots, Y_n$ is presented as:

$$g(y_1, y_2, \ldots, y_n) = n! f(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad -\infty < y_1 < y_2 < \cdots < y_n < \infty.$$

We can also derive the joint pdf of any two order statistics, if h and k are two integers such that $1 \le h < k \le n$, the joint density of Y_h and Y_k is given by

$$g_{h,k}(y_h, y_k) = \frac{n!}{(h-1)!(k-h-1)!(n-k)!} [F(y_h)]^{h-1} [F(y_k) - F(y_h)]^{k-h-1} [1 - F(y_k)]^{n-k} f(y_h) f(y_k),$$

$$-\infty < x_h < x_k < \infty$$

EXAMPLE 5.13

If $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a r.s of size n from $N(\mu, \sigma^2)$. Find the joint pdf of $Y_1, Y_2, \ldots, Y_n$.

Solution. The joint pdf of the order statistics $Y_1, Y_2, \ldots, Y_n$ is

$$g(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f(y_i) = n! (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2}, \quad -\infty < y_1 < y_2 < \cdots < y_n.$$

◀

EXAMPLE 5.14

Let X_1, X_2, X_3 be the order statistics of a r.s of size 3 from $U(0,1)$.

1. Find the joint pdf of Y_1 and Y_3.
2. Find the marginal pdf of $Z = Y_3 - Y_1$.

Solution. Since X_1, X_2, X_3 be the order statistics of a r.s of size 3 from $U(0,1)$, then: $f(x) = 1, 0 < x < 1$. and

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

1. In order to get the joint pdf of Y_1 and Y_3, we have $n = 3$, $h = 1$ and $k = 3$. Therefore,

$$g_{1,3}(y_1, y_3) = \frac{3!}{0!1!0!}[F(y_1)]^0[F(y_3) - F(y_1)]^1[1 - F(y_3)]^0 f(y_1)f(y_3) = 6(y_3 - y_1), \quad 0 < y_1 < y_3 < 1.$$

2. We have the transformation $Z = Y_3 - Y_1$, set $W = Y_3$. Now, the one-to-one functions $z = y_3 - y_1$, $w = y_3$ map the space $\mathcal{A} = \{(y_1, y_3) : 0 < y_1 < y_3 < 1\}$ onto $\mathcal{B} = \{(z, w) : 0 < z < w < 1\}$, with inverses $y_1 = w - z$ and $y_3 = w$, and

$$J = \frac{\partial(y_1, y_3)}{\partial(z, w)} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1.$$

Then, the joint pdf of Z and W is

$$h(z, w) = g_{1,3}(w - z, w)|J| = 6(w - (w - z)) = 6z, \quad 0 < z < w < 1.$$

The marginal pdf of Z is

$$h_1(z) = \int_w h(z, w)dw = \int_z^1 6z dw = 6z\Big[w\Big]_z^1 = 6z(1 - z), \quad 0 < z < 1.$$

◀

CHAPTER 6

SAMPLING DISTRIBUTIONS

6.1 Introduction

This chapter helps make the transition between probability and inferential statistics. Given a sample of n observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters). Prior to obtaining data, there is uncertainty as to which of all possible samples will occur. Because of that, estimates such as $\bar{x}$, $\tilde{x}$ and s will vary from one sample to another. The behaviour of such estimates in repeated sampling is described by what are called sampling distributions. Any particular sampling distribution will give an indication of how close the estimate is likely to be to the value of the parameter being estimated. We will start with some important definition and theorems that are needed in what follows.

Recall the definitions from Chapter‘5, which are

Definition*:** A function of one or more r.v's which does not depend on any unknown parameter is called a ***Statistic. For example, $Y = \sum X_i$ is a statistic, while $Y = X - \mu/\sigma$ is not a statistic unless μ and σ is known. A statistic is any quantity whose value can be calculated from sample data. Therefore, a statistic is a random variable and will be denoted by an upper case letter; a lower case letter is used to represent the calculated or observed value of the statistic.

***Definitions*:** Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from any distribution, we define

1. The statistic $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the sample mean.

2. The statistic $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(X_i^2 - n\bar{X}\right)\right]$ is called the sample variance.

Please enter \offprintinfo{(Title, Edition)}{(Author)}
at the beginning of your document.

EXAMPLE 6.1

The time that takes to serve a customer at the cash register in a mini-market is a r.v having an exponential distribution with parameter $1/\lambda$. Suppose X_1 and X_2 are service times for two different customers, assumed independent of each other. Consider the total service time $T_o = X_1 + X_2$ for the two customers, also a statistic. The cdf of T_o is, for $t \geq 0$,

$$F_{T_o}(t) = \Pr(T_o \leq t) = \Pr(X_1 + X_2 \leq t) = \iint\limits_{\{(x_1,x_2):x_1+x_2\leq t\}} f(x_1, x_2)dx_1dx_2$$
$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} dx_2 dx_1 = \int_0^t (\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t})dx_1 = 1 - e^{-\lambda t} - \lambda t e^{-\lambda x_1}$$

The pdf of T_o is obtained by differentiating $F_{T_o}(t)$:

$$f_{T_o}(t) = \lambda^2 t e^{-\lambda t}, \quad t \geq 0. \tag{6.1}$$

This is a gamma pdf ($\alpha = 2$ and $\beta = 1/\lambda$).

The pdf of $\bar{X} = T_o/2$ can be obtained by the cdf method of Section as

$$f_{\bar{X}}(\bar{x}) = 4\lambda^2 \bar{x} e^{-2\lambda \bar{x}}, \quad \bar{x} \geq 0. \tag{6.2}$$

The mean and variance of the underlying exponential distribution are $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. Using expressions (6.1) and (6.2), it can be verified that $E[\bar{X}] = 1/\lambda$, $Var(\bar{X}) = 1/(2\lambda^2)$, $E[T_o] = 2/\lambda$ and $Var(T_o) = 2/\lambda^2$. These results again suggest some general relationships between means and variances of $\bar{X}$, T_o, and the underlying distribution.

6.2 Sampling Distributions Related to the Normal Distribution

We have already noted that many phenomena observed in the real world have relative frequency distributions that can be modelled adequately by a normal probability distribution. Thus, in many applied problems, it is reasonable to assume that the observable random variables in a random sample, $X_1, X_2, \ldots, X_n$, are independent with the same normal density function. We present it formally in the following theorem.

***Theorem 1*:**

Let $X_1, X_2, \ldots, X_n$ be independent r.v's with mgf of X_i, $i = 1, 2, \ldots, n$ is $M_{X_i}(t_i)$. Then the r.v $Y = \sum_{i=1}^n k_i X_i$ has an mgf $M_Y(t) = \prod_{i=1}^n M_{X_i}(k_i t)$.

***Proof*:**

In order to evaluate the mgf of $Y = \sum_{i=1}^n k_i X_i$, we consider

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\sum_{i=1}^n k_i X_i}\right] = E\left[\prod_{i=1}^n e^{tk_i X_i}\right]$$
$$= \prod_{i=1}^n E\left[e^{tk_i X_i}\right] = \prod_{i=1}^n M_{X_i}(k_i t).$$
$$.. \tag{6.3}$$

***Theorem 2*:**

Let $X_1, X_2, \ldots, X_n$ be independent r.v's where $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. Then the r.v $Y = \sum_{i=1}^{n} k_i X_i$ has a normal distribution, $N\left(\sum_{i=1}^{n} k_i\mu_i, \sum_{i=1}^{n} k_i^2\sigma_i^2\right)$.

***Proof*:**

Since each $X_i \sim N(\mu_i, \sigma_i^2)$, then the mgf of any X_i is $M_{X_i}(t_i) = e^{\mu_i t_i + \frac{1}{2}\sigma_i^2 t_i^2}$, for each $i = 1, 2, \ldots, n$.

If $M_Y(t)$ is the mgf of Y, the according to theorem 1,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(k_i t) = \prod_{i=1}^{n} e^{\mu_i k_i t + \frac{1}{2}\sigma_i^2 k_i^2 t^2} = e^{(\sum_{i=1}^{n} k_i\mu_i)t + \frac{1}{2}(\sum_{i=1}^{n} k_i^2\sigma_i^2)t^2},$$

Which is the mgf of the r.v $Y \sim N(\sum_{i=1}^{n} k_i\mu_i, \sum_{i=1}^{n} k_i^2\sigma_i^2)$

***Note*:** If $X_1, X_2, \ldots, X_n$ from theorem 2 represent a r.s of size n from $N(\mu, \sigma^2)$, then the r.v

$$Y = \sum_{i=1}^{n} k_i X_i \sim N\left(\mu\sum_{i=1}^{n} k_i, \sigma^2\sum_{i=1}^{n} k_i^2\right)$$

.

***Theorem 3*: (The Distribution of $\bar{X}$)**

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is normally distributed with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma_{\bar{X}}^2 = \sigma^2/n$, i.e $\bar{X} \sim N(\mu, \sigma^2/n)$.

***Proof*:**

Since $X_i \sim N(\mu, \sigma^2)$, and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{1}{n} X_i$. Then, according Theorem 2, taking $k_i = 1/2$,

$$\bar{X} \sim N\left(\sum_{i=1}^{n} \frac{1}{n}\mu, \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2\right) \equiv N\left(\mu, \sigma^2/n\right).$$

EXAMPLE 6.2

$X_1, X_2, \ldots, X_{25}$ be a r.s of size 25 from $N(75, 100)$. Find $\Pr(71 < \bar{X} < 79)$.

Solution. We have $n = 25$, $\mu = 75$ and $\sigma^2 = 100$, and since each $X_i \sim N(75, 100)$, then according Theorem 2

$$\bar{X} \sim N(70, 100/25) \equiv N(75, 4).$$

Therefore,

$$\begin{aligned}\Pr(71 < \bar{X} < 79) &= \Pr\left(\frac{71-75}{2} < \frac{\bar{X}-75}{2} < \frac{79-75}{2}\right) = \Pr(-2 < Y < 2)\\ &= 2\Pr(Y < 2) - 1 = 2(0.977) - 1 = 0.954\end{aligned}$$

◀

EXAMPLE 6.3

A bottling machine can be regulated so that it discharges an average of μ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day (all bottled with the same machine setting), and the ounces of fill are measured for each. Find the probability that the sample mean will be within 0.3 ounce of the true mean μ for the chosen machine setting.

Solution. If $X_1, X_2, \ldots, X_9$ denote the ounces of fill to be observed, then we know that the X_i's are normally distributed with mean μ and variance $\sigma^2 = 1$ for $i = 1, 2, \ldots, 9$. Therefore, by Theorem 2, $\bar{X}$ possesses a normal sampling distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n = 1/9$. We want to find

$$\Pr(|\bar{X} - \mu| \leq 0.3) = \Pr(-0.3 \leq \bar{X} - \mu \leq 0.3) = \Pr\left(-\frac{0.3}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right)$$
$$= \Pr\left(-\frac{0.3}{1/\sqrt{9}} \leq Z \leq \frac{0.3}{1/\sqrt{9}}\right) = \Pr(-0.9 \leq Z \leq 0.9) = 2(0.1841) - 1 = 0.6318.$$

Thus, the probability is only 0.6318 that the sample mean will be within 0.3 ounce of the true population mean. ◀

EXAMPLE 6.4

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed r.v with $\mu = 1.5$ min and $\sigma = 0.35$ min. Suppose five rats are selected. Let $X_1, \ldots, X_5$ denote their times in the maze. Assuming the X_i's to be a random sample from this normal distribution, what is the probability that the total time $T_o = X_1 + \cdots + X_5$ for the five is between 6 and 8 min? Determination of the probability that the sample average time X is at most 2.0 min.

Solution. By Theorem 2, T_o has a normal distribution with $\mu_o = n\mu = 5(1.5) = 7.5$ and variance $\sigma_o^2 = n\sigma^2 = 5(0.1225) = 0.6125$, so $\sigma_o = 0.783$. To standardise T_o, subtract μ_o and divide by σ_o:

$$\Pr(6 \leq T_o \leq 8) = \Pr\left(\frac{6 - 7.5}{0.783} \leq Z \leq \frac{8 - 7.5}{0.783}\right) = \Pr(-1.92 \leq Z \leq 0.64) = 0.7115$$

The determination of the probability that the sample average time X (a normally distributed variable) is at most 2 min, we require $\mu_{\bar{X}} = \mu = 1.5$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.35/\sqrt{5} = 0.1565$. Then

$$\Pr(\bar{X} \leq 2) = \Pr\left(Z \leq \frac{2 - 1.5}{0.1565}\right) = \Pr(Z \leq 3.19) = 0.9993.$$

◀

Theorem 4

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from a normal distribution with mean μ and variance σ^2. Then $Z_i = (X_i - \mu)/\sigma$ are independent, standard normal random variables, $i = 1, 2, \ldots, n$, and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

has a χ^2 distribution with n degrees of freedom (df).

EXAMPLE 6.5

If $Z_1, Z_2, \ldots, Z_6$ denotes a r.s from the standard normal distribution, find a number b such that

$$\Pr\left(\sum_{i=1}^{n} Z_i^2 \leq b\right) = 0.95.$$

Solution. By Theorem 4,$\sum_{i=1}^{6} Z_i^2$ has a χ^2 distribution with 6 df. Looking at the tables, in the row headed 6 df and the column headed $\chi^2_{0.05}$, we see the number 12.5916. Thus,

$$\Pr\left(\sum_{i=1}^{6} Z_i^2 > 12.9516\right) = 0.05 \equiv \Pr\left(\sum_{i=1}^{6} Z_i^2 \leq 12.9516\right) = 0.95,$$

and $b = 12.5916$ is the 0.95 quantile (95th percentile) of the sum of the squares of six independent standard normal random variables. ◀

***Theorem 5*: (The Distribution of $\frac{(n-1)S^2}{\sigma^2}$)**

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

has a χ^2 distribution with $(n-1)$ df. Also, $\bar{X}$ and S^2 are independent r.v's, i.e, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

EXAMPLE 6.6

Let $X_1, X_2, \ldots, X_6$ be a r.s of size 6 from $N(\mu, 10)$. Find $\Pr(2.3 < S^2 < 22.2)$

Solution. Since $X_i \sim N(\mu, 10)$, for $i = 1, 2, \ldots, 6$. Then, according Theorem 5, $\frac{(n-1)S^2}{\sigma^2} = \frac{5S^2}{10} = \frac{S^2}{2} \sim \chi^2(5)$. Therefore,

$$\begin{aligned}\Pr(2.3 < S^2 < 22.2) &= \Pr\left(\frac{2.3}{2} < \frac{S^2}{2} < \frac{22.2}{2}\right) = \Pr(1.15 < Y < 11.1)\\ &= \Pr(Y < 11.1) - Pr(Y < 1.15) = 0.95 - 0.05 = 0.9.\end{aligned}$$

◀

6.3 The Central Limit Theorem

If we sample from a normal population, Theorem 3 tells us that $\bar{X}$ has a normal sampling distribution. But what can we say about the sampling distribution of $\bar{X}$ if the variables X_i are not normally distributed? Fortunately, $\bar{X}$ will have a sampling distribution that is approximately normal if the sample size is large. The formal statement of this result is called the central limit theorem.

***Theorem*: (The Central Limit Theorem)**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit as $n \to \infty$, the standardised versions of $\bar{X}$ and T_o have the standard normal distribution. That is,

$$\lim_{n\to\infty} \Pr\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Pr(Z \leq z)$$

or

$$\lim_{n\to\infty} \Pr\left(\frac{T_o - n\mu}{\sigma\sqrt{n}} \le z\right) = \Pr(Z \le z)$$

on another words,

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{T_o - n\mu}{\sigma\sqrt{n}} \sim N(0,1).$$

EXAMPLE 6.7

Let $X_1, X_2, \ldots, X_n$ be a r.s of size $n = 75$ from $U(0,1)$. Approximate $\Pr(0.45 < \bar{X} < 0.55)$, where $\bar{X}$ is the sample mean.

Solution. Since $X \sim U(0,1)$, then $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$. Then according to the central limit theorem, the r.v $Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has a limiting $N(0,1)$, that is

$$Y = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{75}(\bar{X} - \mu)}{\sqrt{1/12}} = \sqrt{(75)(12)}(\bar{X} - 0.5) = 30(\bar{X} - 0.5).$$

The approximate value of

$$\begin{aligned}\Pr(0.45 < \bar{X} < 0.55) &= \Pr[30(0.45 - 0.5) < 30(\bar{X} - 0.5) < 30(0.55 - 0.5)] \\ &= \Pr(-1.5 < Y < 1.5) = 2\Pr(Y \le 1.5) - 1 \\ &= 2(0.933) - 1 = 0.866.\end{aligned}$$

◀

EXAMPLE 6.8

The service times for customers coming through a checkout counter in a retail store are independent random variables with mean 1.5 minutes and variance 1.0. Approximate the probability that 100 customers can be served in less than 2 hours of total service time.

Solution. If we let X_i denote the service time for the ith customer, then we want

$$\Pr\left(\sum_{i=1}^{100} X_i \le 120\right) = \Pr\left(\bar{X} \le \frac{120}{100}\right) = \Pr(\bar{X} \le 1.2).$$

Because the sample size is large, the central limit theorem tells us that $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}} = \mu = 1.5$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n = 1/100$. Therefore, using the tables, we have

$$\begin{aligned}\Pr(\bar{X} \le 1.2) &= \Pr\left(\frac{\bar{X} - 1.5}{1/\sqrt{100}} \le \frac{1.2 - 1.5}{1/\sqrt{100}}\right) \\ &\approx \Pr[Z \le (1.2 - 1.5)10] = \Pr(Z \le -3) = 0.0013.\end{aligned}$$

Thus, the probability that 100 customers can be served in less than 2 hours is approximately 0.0013. This small probability indicates that it is virtually impossible to serve 100 customers in only 2 hours. ◀

CHAPTER 7

POINT ESTIMATION

7.1 Introduction

Given a parameter of interest, such as a population mean μ or population proportion p, the objective of point estimation is to use a sample to compute a number that represents in some sense a good guess for the true value of the parameter. The resulting number is called a *point estimate*. In Section 7.1, we present some general concepts of point estimation. In Section 7.2, we describe and illustrate two important methods for obtaining point estimates: the method of moments and the method of maximum likelihood.

Obtaining a point estimate entails calculating the value of a statistic such as the sample mean $\bar{X}$ or sample standard deviation S. We should therefore be concerned that the chosen statistic contains all the relevant information about the parameter of interest. The idea of no information loss is made precise by the concept of sufficiency.

7.2 General Concepts and Criteria

Statistical inference is frequently directed toward drawing some type of conclusion about one or more parameters (population characteristics). To do so requires that an investigator obtain sample data from each of the populations under study. Conclusions can then be based on the computed values of various sample quantities.

***Definition*: A Point Estimator** of a parameter θ is a single number that can be regarded as a sensible value for θ. A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ.

***Notes*:**

1. we will denote to the estimate of an unknown parameter θ by $\hat{\theta}$.

Please enter \offprintinfo{(Title, Edition)}{(Author)}
at the beginning of your document.

2. Many different estimators (rules for estimating) may be obtained for the same population parameter.
3. The sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is one possible point estimator of the population mean μ.

EXAMPLE 7.1

An automotive manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models. Let $X =$ the number of crashes that result in no visible damage to the automotive. The parameter to be estimated is $p =$ the proportion of all such crashes that result in no damage [alternatively, $p = \Pr$(no damage in a single crash)]. If X is observed to be $x = 15$, the most reasonable estimator and estimate are

$$\text{estimator } \hat{p} = \frac{X}{n} \qquad \text{estimate } = \frac{x}{n} = \frac{15}{25} = 0.6$$

If for each parameter of interest there were only one reasonable point estimator, there would not be much to point estimation. In most problems, though, there will be more than one reasonable estimator.

7.2.1 The Bias and Mean Square Error of Point Estimators

Suppose we have two measuring instruments; one instrument has been accurately calibrated, but the other systematically gives readings smaller than the true value being measured. When each instrument is used repeatedly on the same object, because of measurement error, the observed measurements will not be identical. However, the measurements produced by the first instrument will be distributed about the true value in such a way that on average this instrument measures what it purports to measure, so it is called an unbiased instrument. The second instrument yields observations that have a systematic error component or bias. In other words, we would like the mean or expected value of the distribution of estimates to equal the parameter estimated; that is, $E[\hat{\theta}] = \theta$. Point estimators that satisfy this property are said to be *unbiased*.

***Definition*:** Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an **unbiased estimator** if $E[\hat{\theta}] = \theta$. If $E[\hat{\theta}] \neq \theta$, $\hat{\theta}$ is said to be biased. the difference $B(\hat{\theta}) = E[\hat{\theta}] - \theta$ is called the bias of $\hat{\theta}$.

In the best of all possible worlds, we could find an estimator $\hat{\theta}$ for which $\hat{\theta} = \theta$ always. However, $\hat{\theta}$ is a function of the sample X_i's, so it is a random variable. For some samples, $\hat{\theta}$ will yield a value larger than θ, whereas for other samples $\hat{\theta}$ will underestimate θ. If we write

$$\hat{\theta} = \theta + \text{error of estimation}$$

then an accurate estimator would be one resulting in small estimation errors, so that estimated values will be near the true value.

A popular way to quantify the idea of $\hat{\theta}$ being close to θ is to consider the squared error $(\hat{\theta} - \theta)^2$. An omnibus measure of accuracy is the mean squared error (expected squared error), which entails averaging the squared error over all possible samples and resulting estimates.

***Definition*:** The **mean square error** of a point estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

The mean square error of an estimator $\hat{\theta}$, $MSE(\hat{\theta})$, is a function of both its variance and its bias. If $B(\hat{\theta})$ denotes the bias of the estimator $\hat{\theta}$, it can be shown that, since

$$Var(X) = E[X^2] - (E[X])^2 \;\Rightarrow\; E[X^2] = Var(X) + (E[X])^2,$$

then,

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [B(\hat{\theta})]^2.$$

EXAMPLE 7.2

Consider once again estimating a population proportion of "successes" p. The natural estimator of p is the sample proportion of successes $\hat{p} = X/n$. The number of successes X in the sample has a binomial distribution with parameters n and p, so $E[X] = np$ and $Var(X) = np(1-p)$. The expected value of the estimator is

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = \frac{1}{n}np = p.$$

Thus the bias of $\hat{p}$ is $p - p = 0$, and therefore $\hat{p}$ is an unbiased estimator. Giving the mean squared error as

$$E[(\hat{p}-p)^2] = Var(\hat{p}) + 0^2 = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{p(1-p)}{n}.$$

Now consider the alternative estimator $\hat{p} = (X+2)/(n+4)$. That is, add two successes and two failures to the sample and then calculate the sample proportion of successes. The bias of the alternative estimator is

$$E\left[\frac{X+2}{n+4}\right] - p = \frac{1}{n+4}E[X+2] - p = \frac{np+2}{n+4} - p = \frac{2/n - 4p/n}{1+4/n}.$$

This bias is not zero unless $p = 0.5$. The variance of the estimator is

$$Var\left(\frac{X+2}{n+4}\right) = \frac{1}{(n+4)^2}Var(X+2) = \frac{Var(X)}{(n+4)^2} = \frac{np(1-p)}{(n+4)^2} = \frac{p(1-p)}{n+8+16/n}.$$

This variance approaches zero as the sample size increases. The mean squared error of the alternative estimator is

$$MSE(\hat{p}) = \frac{p(1-p)}{n+8+16/n} + \left(\frac{2/n - 4p/n}{1+4/n}\right)^2.$$

7.2.2 Some Common Unbiased Point Estimators

In this section, we focus on some estimators that merit consideration on the basis of intuition. For example, it seems natural to use the sample mean $\bar{X}$ to estimate the population mean μ and S^2 as the sample variance unbiased estimator of σ^2. Also to use the sample proportion $\hat{p} = x/n$ to estimate a binomial parameter p.

***Propositions*:**

1. When X is a binomial r.v with parameters n and p, the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p.
2. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ, then $\bar{X}$ is an unbiased estimator of μ.

EXAMPLE 7.3

Let $X_1, X_2, \ldots, X_n$ be a random sample with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Show that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is an unbiased estimator for σ^2.

Solution. It can be shown that

$$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2.$$

Hence

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=E\left[\sum_{i=1}^{n}X_i^2\right]-nE[\bar{X}^2].$$

Notice that $E[X_i^2]$ is the same for $i=1,2,\ldots,n$. We use this and the fact that the variance of a r.v is given by $Var(X)=E[X^2]-(E[X])^2$ to conclude that $E[X_i^2]=Var(X_i)+(E[X_i])^2=\sigma^2+\mu^2$, $E[\bar{X}^2]=Var(\bar{X})+(E[\bar{X}])^2=\sigma^2/n+\mu^2$, and that

$$\begin{aligned}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]&=\sum_{i=1}^{n}(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right)=n(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\\&=n\sigma^2-\sigma^2=(n-1)\sigma^2.\end{aligned}$$

It follows that

$$E[S^2]=\frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{1}{n-1}(n-1)\sigma^2=\sigma^2,$$

so we see that S^2 is an unbiased estimator for σ^2. ◀

***Proposition*:** Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ Is called the **minimum variance unbiased estimator** (MVUE) of θ. Since $MSE=variance+(bias)^2$, seeking an unbiased estimator with minimum variance is the same as seeking an unbiased estimator that has minimum mean squared error.

***Definition*:** The standard error of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}}=\sqrt{Var(\hat{\theta})}$. If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $s_{\hat{\theta}}$ yields the estimated standard error (estimated standard deviation) of the estimator. The estimated standard error can be denoted either by $\hat{\sigma}_{\hat{\theta}}$ (the ˆ over σ emphasises that $\sigma_{\hat{\theta}}$ is being estimated) or by $S_{\hat{\theta}}$.

EXAMPLE 7.4

Assuming that breakdown voltage is normally distributed, $\hat{\mu}=\bar{X}$ is the best estimator of μ. If the value of σ is known to be 1.5, the standard error of $\bar{X}$ is $\sigma_{\bar{X}}=\sigma/\sqrt{n}=1.5/\sqrt{20}=0.335$. If, as is usually the case, the value of σ is unknown, the estimate $\hat{\sigma}=s=1.462$ is substituted into $\sigma_{\bar{X}}$ to obtain the estimated standard error $\hat{\sigma}_{\bar{X}}=s_{\bar{X}}=s/\sqrt{n}=1.462/\sqrt{20}=0.327$.

EXAMPLE 7.5

Back to Example 7.1, the standard error of $\hat{p}=X/n$ is

$$\sigma_{\hat{p}}=\sqrt{Var(X/n)}=\sqrt{\frac{Var(X)}{n^2}}=\sqrt{\frac{npq}{n^2}}=\sqrt{\frac{pq}{n}}.$$

Since p and $q=1-p$ are unknown (else why estimate?), we substitute $\hat{p}=x/n$ and $\hat{q}=1-x/n$ into $\sigma_{\hat{p}}$, yielding the estimated standard error $\hat{\sigma}_{\hat{p}}=\sqrt{\hat{p}\hat{q}/n}=\sqrt{(0.6)(0.4)/25}=0.098$. Alternatively, since the largest value of pq is attained when $p=q=0.5$, an upper bound on the standard error is $\sqrt{1/(4n)}=0.1$.

7.3 Methods of Point Estimation

So far the point estimators we have introduced were obtained via intuition and/or educated guesswork. We now discuss two "constructive" methods for obtaining point estimators: the method of moments and the method of maximum likelihood. By constructive we mean that the general definition of each type of estimator suggests explicitly how to obtain the estimator in any specific problem. Although maximum likelihood estimators are generally preferable to moment estimators because of certain efficiency properties, they often require significantly more computation than do moment estimators. It is sometimes the case that these methods yield unbiased estimators.

7.3.1 The method of Moment

The basic idea of this method is to equate certain sample characteristics, such as the mean, to the corresponding population expected values. Then solving these equations for unknown parameter values yields the estimators.

***Definition*:** Let $X_1, X_2, \ldots, X_n$ be a r.s from pdf $f(x)$. For $k = 1, 2, 3, \ldots$, the kth population moment, or kth moment of the distribution $f(x)$, is $\mu_r = E[X^k]$. The kth sample moment is $M_r = (1/n)\sum_{i=1}^{n} X_i^k$.

Thus the first population moment is $E[X] = \mu$ and the first sample moment is $\sum X_i/n = \bar{X}$. The second population and sample moments are $E[X^2]$ and $\sum X_i^2/n$, respectively. The population moments will be functions of any unknown parameters $\theta_1, \theta_2, \ldots$.

***Definition*:** Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x; \theta_1, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. Then **the moment estimators** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are obtained by equating the first m sample moments to the corresponding first m population moments and solving for $\theta_1, \ldots, \theta_m$.

EXAMPLE 7.6

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from $P(\lambda)$. Estimate λ by the method of moment.

Solution. Since $X_i \sim P(\lambda)$, then $f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$. In this case, we have one unknown parameter, λ, so we set

$$\mu_r = M_r \text{ at } \lambda = \hat{\lambda}, r = 1$$

therefore,

$$\mu_1 = E[X] = \lambda \text{ and } M_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

So, $\hat{\lambda} = \bar{x}$ is the moment estimate of λ. ◀

EXAMPLE 7.7

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from $Exp(1/\theta)$. Estimate λ by the method of moment.

Solution. Since $X \sim Exp(1/\theta)$, then $f(x) = \theta e^{-\theta x}$, $0 < x < \infty$. In this case, we have one unknown parameter, λ, so we set

$$\bar{X} = \frac{1}{\theta} \Rightarrow \hat{\theta} = \frac{1}{\bar{x}}.$$

◀

EXAMPLE 7.8

Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution with parameters α and β. Find the Moment estimators for α and β.

Solution. Since $X_i \sim G(\alpha, \beta)$, then $\mu = E[X] = \alpha\beta$ and $E[X^2] = \beta^2\Gamma(\alpha+2)/\Gamma(\alpha) = \beta^2(\alpha+1)\alpha$. Then The moment estimators of α and β are obtained by solving

$$\bar{X} = \alpha\beta, \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \alpha(\alpha+1)\beta^2.$$

Since $\alpha(\alpha+1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$ and the first equation implies $\alpha^2\beta^2 = (\bar{X})^2$, the second equation becomes

$$\frac{1}{n}\sum X_i^2 = (\bar{X})^2 + \alpha\beta^2.$$

Now dividing each side of this second equation by the corresponding side of the first equation and substituting back gives the estimators

$$\hat{\alpha} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} = \frac{n(\bar{X})^2}{(n-1)S^2}, \quad \hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - (\bar{X})^2}{\bar{X}} = \frac{(n-1)S^2}{n(\bar{X})^2}$$

◀

EXAMPLE 7.9

Let $X_1, \ldots, X_n$ be a r.s from a generalized negative binomial distribution with parameters r and p. Estimate the value of r and p.

Solution. Since $E[X] = r(1-p)/p$ and $Var(X) = r(1-p)/p^2$, $E[X^2] = Var(X) + (E[X])^2 = r(1-p)(r - rp + 1)/p^2$. Equating

$$E[X] = \bar{X} \quad \text{and} \quad E[X^2] = (1/n)\sum X_i^2$$

eventually gives,

$$\hat{p} = \frac{\bar{X}}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} = \frac{n\bar{X}}{(n-1)S^2}, \quad \hat{r} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2 - \bar{X}} = \frac{n(\bar{X})^2}{(n-1)S^2 - \bar{X}}.$$

◀

7.3.2 Maximum Likelihood Method

The method of maximum likelihood was first introduced by R. A. Fisher, a geneticist and statistician, in the 1920s. Most statisticians recommend this method, at least when the sample size is large, since the resulting estimators have certain desirable efficiency properties

***Definition*:** The **Likelihood Function** of a r.s $X_1, X_2, \ldots, X_n$ of size n from a distribution with a pdf $f(x)$ with parameters $\theta_1, \theta_2, \ldots, \theta_m$, is defined to be the joint pdf of the n r.v's $X_1, X_2, \ldots, X_n$ which is considered as a function of θ's and denoted by $L(\tilde{\theta}; \tilde{x})$, i.e

$$L(\tilde{\theta}; \tilde{x}) = L(\theta_1, \theta_2, \ldots, \theta_m; x_1, x_2, \ldots, x_n) = f(\tilde{x}; \tilde{\theta}) = \prod_{i=1}^{n} f(x_i).$$

The maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ are those values of the θ_i's that maximize the likelihood function, so that

$$L(x_1, x_2, \ldots, x_n; \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m) \geq L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m), \ \ \forall \theta_1, \theta_2, \ldots, \theta_m$$

When the X_i's are substituted in place of the x_i's, the maximum likelihood estimators (mle's) result.

***Notes*:**

1. Many likelihood functions satisfy the condition that mle is the solution of the likelihood equations

$$\frac{\partial L(\tilde{\theta}; \tilde{x})}{\partial \theta_r} = 0, \ \text{ at } \tilde{\theta} = \hat{\tilde{\theta}}, \ \ r = 1, 2, \ldots, m.$$

2. Since $L(\tilde{\theta}; \tilde{x})$ and $\ln L(\tilde{\theta}; \tilde{x})$ have their maxima at the same value of $\tilde{\theta}$, so it is some times easier to find the maximum of the logarithm of the likelihood function. In this case, the mle $\tilde{\theta}$ of θ which maximises $L(\tilde{\theta}; \tilde{x})$ may be given by the solution of

$$\frac{\partial \ln L(\tilde{\theta}; \tilde{x})}{\partial \theta_r} = 0, \ \text{ at } \tilde{\theta} = \hat{\tilde{\theta}}, \ \ r = 1, 2, \ldots, m.$$

EXAMPLE 7.10

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from Bernoulli distribution with a parameter p ($X \sim Ber(p)$). Estimate p using the maximum likelihood method.

Solution. Since $X \sim Ber(p)$, then the pdf of X is: $f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$. The likelihood function is

$$L(p; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

It is easier to maximise $\ln L(p, x_1, \ldots, x_n)$, and

$$\ln l(p; \tilde{x}) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p),$$

Hence,

$$\frac{\partial \ln L}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\frac{\partial \ln L}{\partial p}\Big|_{p=\hat{p}} = 0 \Rightarrow \frac{\sum_{i=1}^{n} x_i}{\hat{p}} - \frac{n - \sum_{i=1}^{n} x_i}{1-\hat{p}} = 0 \Rightarrow \frac{n\bar{x}}{\hat{p}} = \frac{n - n\bar{x}}{1-\hat{p}}$$

◀

Therefore,

$$n\bar{x} - n\bar{x}\hat{p} = n\hat{p} - n\bar{x}\hat{p} \Rightarrow \hat{p} = \bar{x}$$

Then, $\bar{X}$ is the mle of p.

EXAMPLE 7.11

Let $X_1, X_2, \ldots, X_n$ be a r.s from a normal distribution with mean μ and variance σ^2. Find the MLEs of μ and σ^2.

Solution. Since, $X \sim N(\mu, \sigma^2)$, then $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$. The likelihood function is

$$L(\nu, \sigma^2; \tilde{x}) = f(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right].$$

Further,

$$\ln L(\mu, \sigma^2; \tilde{x}) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

The MLEs of μ and σ^2 are the values that make $\ln L(\mu, \sigma^2; \tilde{x})$ a maximum. Taking derivatives with respect to μ and σ^2, we obtain

$$\frac{\partial \ln L(\mu, \sigma^2; \tilde{x})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu),$$

and

$$\frac{\partial \ln L(\mu, \sigma^2; \tilde{x})}{\partial \sigma} = -\left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Setting these derivatives equal to zero and solving simultaneously, we obtain from the first equation

$$\frac{1}{\hat{\sigma}}^2 \sum_{i=1}^{n}(x_i - \hat{\mu}) = 0, \quad \text{or} \quad \sum_{i=1}^{n} x_i - n\hat{\mu} = 0, \quad \text{and} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Substituting $\bar{x}$ for $\hat{\mu}$ in the second equation and solving for $\hat{\sigma}^2$, we have

$$-\left(\frac{n}{\hat{\sigma}^2}\right) + \frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0, \quad \text{or} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

◀

EXAMPLE 7.12

Let $X_1, X_2, \ldots, X_n$ be a r.s of size n from Gamma distribution with parameters α and β. Find the MLEs α and β.

Solution. The pdf of a r.v $X \sim G(\alpha, \beta)$ is: $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $0 < x < \infty$. The likelihood function is

$$L(\aleph, \beta; \tilde{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta} = [\Gamma(\alpha)]^{-n}\beta^{-n\alpha}\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\frac{1}{\beta}\sum_{i=1}^{n} x_i},$$

Also,

$$\ln L(\alpha, \beta; \tilde{x}) = -n\ln\Gamma(\alpha) - n\alpha\ln\beta + (\alpha-1)\sum_{i=1}^{n}\ln x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i.$$

Now, to maximise $\ln L(\alpha, \beta; \tilde{x})$,

$$\frac{\partial \ln L}{\partial \alpha} = -n\psi(\alpha) - n \ln \beta + \sum_{i=1}^{n} \ln x_i,$$

where, $\psi(\alpha) = \frac{d}{d\alpha} \ln \Gamma(\alpha)$. Also

$$\frac{\partial \ln L(\alpha, \beta; \tilde{x})}{\partial \beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i.$$

Set $\frac{\partial \ln L}{\partial \alpha} = 0$ and $\frac{\partial \ln L}{\partial \beta} = 0$ at $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$, we have

$$-n\psi(\hat{\alpha}) - n \ln \hat{\beta} + \sum_{i=1}^{n} \ln x_i = 0 \tag{7.1}$$

and

$$-\frac{n\hat{\alpha}}{\hat{\beta}} + \frac{1}{\hat{\beta}} \sum_{i=1}^{n} x_i = 0 \Rightarrow n\hat{\alpha} = \frac{n\bar{x}}{\hat{\beta}} \Rightarrow \hat{\alpha}\hat{\beta} = \bar{x}. \tag{7.2}$$

From equations (7.1) and (7.2), we get

$$-n\psi(\hat{\alpha}) - \ln \frac{\bar{x}}{\hat{\alpha}} + \sum_{i=1}^{n} \ln x_i = 0 \Rightarrow -n\psi(\hat{\alpha}) - n \ln \bar{x} + n \ln \hat{\alpha} + \sum_{i=1}^{n} \ln x_i,$$

hence

$$n[\ln \hat{\alpha} - \psi(\hat{\alpha})] = n \ln \bar{x} - \sum_{i=1}^{n} \ln x_i \tag{7.3}$$

There is no analytical solution for equation (7.3) because of $\psi(\hat{\alpha})$. In this case, we need to approximate the value of $\hat{\alpha}$ and $\hat{\beta}$ by the use of numerical methods. ◀